ADVANCES IN DETERMINISTIC, STOCHASTIC, AND SEMISTOCHASTIC QUANTUM CHEMISTRY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Adam Anthony Holmes

January 2017

ADVANCES IN DETERMINISTIC, STOCHASTIC, AND SEMISTOCHASTIC QUANTUM CHEMISTRY

Adam Anthony Holmes, Ph.D.

Cornell University 2017

In this dissertation, I present my original research in the development of algorithms for computing ground-state properties of strongly-correlated electronic systems from first principles. I present three main algorithms.

First, I present a *semistochastic* projection algorithm, dubbed Semistochastic Quantum Monte Carlo, which combines the best qualities of deterministic and stochastic methods for projecting out a ground state wavefunction in a basis of Slater determinants. This new algorithm can treat systems as large as a fully-stochastic algorithm can, while dramatically reducing the statistical uncertainty and bias by treating the most important part of the problem deterministically.

Second, I present an efficient algorithm for sampling many-particle states in Fock space with probability proportional to the Hamiltonian matrix element connecting them to a reference state, which I refer to as the *heat-bath* distribution. This sampling algorithm, referred to as Efficient Heat-bath Sampling in Fock Space, factors and approximates the heat-bath probabilities in such a way that they can be efficiently stored and sampled, without having to enumerate all of the possible excitations. Efficient Heat-bath Sampling dramatically improves the efficiency of stochastic Fock space methods by sampling the more relevant Slater determinants more frequently.

Third, I present the deterministic analog of Efficient Heat-bath Sampling, which enables one to generate all Slater determinants that are connected to a reference by

Hamiltonian matrix elements larger in magnitude than a cutoff, without wasting any time on those determinants that do not meet the cutoff. This deterministic heat-bath "sampling" algorithm is then incorporated into a highly-efficient quantum chemistry algorithm that I call Heat-bath Configuration Interaction, which first generates a variational wavefunction and then computes the lowest-order perturbative correction. Both the variational and perturbative stages of Heat-bath Configuration Interation make use of deterministic heat-bath "sampling" to perform highly efficient calculations using only the most important Slater determinants.

BIOGRAPHICAL SKETCH

Adam A. Holmes was born on September 1, 1988, in Huntsville, Alabama. From a young age, he had a strong interest in math, science, and programming.

After graduating early from the University of North Carolina at Chapel Hill with a double major in physics and mathematics, Adam started his Ph.D. program in the physics department at Cornell University in Fall 2010. He has been working with Prof. Cyrus Umrigar on developing and implementing a variety of electronic structure theory algorithms, ranging from fully-stochastic to fully-deterministic to *semistochastic* algorithms combining the best of both worlds.

Adam spent the summer of 2013 doing research with Dr. Miguel Morales at Lawrence Livermore National Laboratory, where he helped implement and run his Semistochastic Quantum Monte Carlo algorithm on some of the largest computers in the world.

After a brief stint as a Quant at Citadel Investment Group in Chicago in 2014, Adam returned to Ithaca with renewed focus and performed his best research yet, developing a framework for efficiently selecting the most important parts of intractably large second-quantized Hamiltonian matrices. Within this framework, Adam developed both a stochastic sampling algorithm, which samples the more relevant electronic configurations more frequently, and a highly efficient deterministic algorithm, which identifies the most important pieces of the Hamiltonian so that limited computing resources are used wisely.

From September to December 2016, Adam will be a Visiting Researcher at Kobe University in Kobe, Japan, applying his efficient heat-bath sampling algorithm to Prof. Seiichiro Ten-no's Model Space Quantum Monte Carlo (MSQMC) algorithm. He hopes that heat-bath sampling will greatly improve the efficiency of MSQMC in its calculation of ground and excited states.

In January 2017, Adam will begin a Staff Scientist position at the Max Planck Institute for Solid State Physics in Stuttgart, Germany, working in the Electronic Structure Theory group of Prof. Ali Alavi. He plans to apply the algorithms he has developed during his Ph.D. to some exciting applications in the field of solar energy.

I dedicate this dissertation to my loving family.

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Prof. Cyrus Umrigar, for his guidance over the past five years. Not only have I learned a great deal from Cyrus's extensive knowledge and experience in quantum Monte Carlo and related algorithms, but his thoroughness and attention to detail have made a strong impression on me. He has been patient and supportive of me as I explored various possible career options, and has helped me find great opportunities over the years, from an internship at Lawrence Livermore National Lab, to my postdoc at the Max Planck Institute, to several international conferences and workshops. Especially over the last year and a half, Cyrus and I have made a great team, often bringing very different (and complementary!) intuitions to the table as we worked together on challenging problems. I also thank him for sharing fresh eggs from his chickens and delicious fruits, vegetables, and herbs from his garden!

I would like to thank the current and former members of the Umrigar group with whom I've had the honor of collaborating: Frank Petruzielo, Hitesh Changlani, Mihir Khadilkar, John Elton, Matt Otten, Mingtong Han, and Junhao Li. I am especially grateful to Hitesh, who not only taught me almost everything I know about Fortran and the Hubbard model, but also took the time to create a perfect first research project for me that enabled me to quickly get up to speed with the rest of the group. I thank him for patiently and thoroughly answering all my questions early on, and I have many fond memories of debates and brainstorming sessions with him.

Cornell has been a terrific place to pursue my extracurricular interests. I am grateful to my friends in the Cayuga Windsurfing Club, the Cornell Ski and Snowboard Club, the Ithaca salsa community, the Cornell Chess Club, the Ithaca kickball and ultimate frisbee leagues, Ithaca Practitioners of Alemaking, and Students

for Justice in Palestine. Windsurfing, kayaking, and even tubing and kneeboarding behind the world-famous *Bateau Perrault* (a 47-year-old, 19-foot power boat that my good friend Alex Bodell and I obtained for free and had a lot of fun fixing up) on Cayuga Lake have been some of my favorite ways to enjoy the beautiful Ithaca summers. Snowboarding trips to Greek Peak, Sugarbush, and Mont Tremblant helped make the long Ithaca winters something to look forward to. The weekly Wednesday salsa nights at Agava are often among the highlights of my week, and I thank all of my dance partners for their patience and readiness to teach me new moves. I have many fond memories in the Cornell Chess Club, learning from Cornell's many amazing players and representing Cornell at countless team tournaments over the years. To my teammates James Critelli, Tom Riccardi, and Robert Xue at the 2013 World Amateur Team Chess Championship: we may have narrowly missed a well-deserved perfect finish, but 2nd place isn't bad! I thank my good friend Peter DelNero for single-handedly starting a new Cornell tradition: the Big Red Pumpkin Regatta. I'll never forget how fun it was paddling across Beebe Lake in a giant pumpkin!

Thank you to all my friends and mentors for so many wonderful experiences over the years and for making Ithaca a home I will truly miss.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The accurate simulation of molecules is profoundly important in many fields. For example, in synthesizing potential molecular structures for artificial photosynthesis, there are many closely-related molecules that may differ from each other in just one or two key transition-metal atoms. Whereas examining all variations experimentally would be very costly (as each molecule would need to be synthesized and measured individually), with an accurate simulation tool, one could very quickly and easily iterate through many variations and calculate the properties of each, in order to identify the best one. Another example is in trying to understand the structure of water. While it would be challenging to experimentally determine the distribution of the number of hydrogen bonds formed by each water molecule in bulk liquid water, an accurate computer simulation could obtain the distribution with a simple for loop. Determining the structure of water is one of the main open problems in physical chemistry.

One would think that computing molecular properties should be a very simple problem. After all, the underlying physical laws governing the interactions of subatomic particles have been known for nearly 90 years; all one must do is apply those laws to the particular molecule in question. Not only that, but most molecular properties and chemical interactions are governed primarily by the interactions of the electrons. Because atomic nuclei are thousands of times more massive than electrons, it is typical to employ the Born-Oppenheimer approximation, namely, the approximation that the nuclei are infinitely-massive point charges, and only the electrons must be treated quantum mechanically. If that wasn't simple enough, the electrons in many molecules move very slowly relative to the speed of light, so

that relativistic effects can either be ignored entirely or treated in a perturbative way.

Thus, the name of the game is simply to solve the non-relativistic Schrodinger equation only for the electrons moving in a field generated by stationary nuclei. The non-relativistic Schrodinger equation is very simple to write down:

$$H\Psi = E\Psi,$$

where the Hamiltonian H contains the fundamental physics of how electrons behave, the wavefunction Ψ describes the state that the electrons are in, and E is the energy of that state. Usually, one is only interested in the properties of the ground state Ψ_0, i.e., the state with lowest energy, and maybe a few low-lying excited states. However, this very simple equation is extremely difficult to solve. In the words of Paul Dirac, "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

What makes this equation so hard to solve is the fact that every electron interacts with every other electron. If there were only one electron, it would be easy. For example, the ground state of the hydrogen atom can be worked out with pencil and paper. However, most interesting molecular systems have many more electrons than one. A hint at the difficulty in solving the many-electron Schrodinger equation can be seen in the difficulty of even describing the wavefunction Ψ. The wavefunction is a function of the coordinates of all the electrons, so even for a small molecule like water, which has only 10 electrons, the wavefunction is a 30-dimensional function. If one were to represent it with a coarse grid of only 10 points in each dimension, there would still be 10^{30} grid points, an enormous amount of

information just to describe a very coarse approximate wavefunction!

A convenient approach to describing wavefunctions is to make use of a basis of single-electron states, or molecular orbitals. The electrons "live" in these orbitals and can transition from one orbital to another. In this paradigm, a many-electron state corresponds to a particular assignment of electrons to orbitals. Because electrons are fermions, these configurations are anti-symmetric with respect to the interchange of a pair of electrons, and are therefore written as determinants of single-particle functions (so-called Slater determinants). An arbitrary many-particle wavefunction in this basis can be written as a linear combination of Slater determinants. While the ground state wavefunction within a basis is not exact, it becomes exact in the limit of an infinitely large basis set, and classes of basis sets have been developed that enable smooth extrapolation to this limit. Of course, the exponential complexity of the problem has not disappeared: the number of Slater determinants one can construct grows combinatorially with the number of electrons and with the number of orbitals. Still, we make use of this basis set approach throughout this thesis.

Because an exact solution (even within a basis set!) is often impossible with our finite computer resources, we seek the best approximate solution to the Schrodinger equation that we can afford. Broadly, there are two main approaches to obtaining approximate solutions: deterministic and stochastic. Deterministic methods often reduce the problem to something computationally tractable by finding the ground state within a reduced Hilbert space, rather than the full, exponentially large, set of Slater determinants. This can be done either by introducing a wavefunction ansatz, as in Density Matrix Renormalization Group (DMRG), or by truncating the set of Slater determinants, as in Selected Configuration Interaction (Selected

CI). One of our main developments in this thesis is a Selected CI algorithm that identifies which determinants to keep or discard iteratively in a way that is nearly optimal on each iteration[1].

Stochastic methods, on the other hand, aspire to sample the exact ground state wavefunction (either in a finite basis or even in an infinite basis!) using random numbers, and then compute expectation values of properties of interest. I use the word "aspire" here, because sampling the exact wavefunction is generally not possible because of the infamous *fermion sign problem*. In the context of basis set wavefunctions, where a highly successful algorithm is called Full Configuration Interaction Quantum Monte Carlo (FCIQMC) [12], the sign problem appears because, when computing the coefficient of a Slater determinant, large positive and negative weights might both arrive on the same determinant on different iterations of the algorithm. While these positive and negative weights would simply cancel in a deterministic algorithm, they would be included in the statistics in the stochastic algorithm, resulting in an exponentially large noise-to-signal ratio. In practice, approximations such as the initiator approximation [17] must be introduced to tame the sign problem, so even the expectation values from stochastic methods are only exact when extrapolated to some limit.

Deterministic and stochastic methods thus both have advantages and disadvantages. Deterministic methods give an exact answer within a reduced problem, without any statistical noise or additional bias due to the sign problem. On the other hand, stochastic methods can work within much larger determinant spaces, by only needing to store sparse "snapshots" of the wavefunction on any iteration,

[1]The final set of determinants will not necessarily be optimal, for two reasons. First, we have not yet explored the optimal "schedule" for how many new determinants to add on each iteration, and the final set of determinants will be sensitive to this schedule. Second, the importance of some determinants may change relative to all the others as new determinants are added, so determinants that are added on early iterations may not be important in the final ground state.

and accumulating statistics from each snapshot before discarding it. Therefore, one may wonder whether a judicious combination of deterministic and stochastic methods could incorporate the best of both worlds. Another of our main developments in this thesis was to do just that, and to combine fully deterministic matrix-vector multiplication in an "important" determinant subspace with the fully deterministic FCIQMC in the rest of the determinant space. We call this Semistochastic FCIQMC (S-FCIQMC), and demonstrate that, when the important subspace is well-chosen, it dramatically reduces both the statistical fluctuations and the initiator bias relative to fully-stochastic FCIQMC [31].

Whether deterministic, stochastic, or semi-stochastic algorithms are used, the (very large) Hamiltonian matrix within a basis set will be the same. Our main contribution in this thesis is to point out that the nonzero Hamiltonian matrix elements span many orders of magnitude and most have a very simple structure, and to introduce an algorithm that takes advantage of these facts to perform efficient calculations using only the important matrix elements. We describe two manifestations of this idea, one stochastic and one deterministic.

First, we demonstrate a stochastic algorithm that enables one to efficiently sample new determinants from the heat-bath distribution, i.e., with probability proportional to the Hamiltonian matrix element connecting the new determinant to a reference determinant [23]. The key contribution here is that we do this efficiently, without having to normalize the probability by summing the Hamiltonian matrix elements corresponding to the many possible new determinants that could be reached from the reference. We call this algorithm Efficient Heat-bath Sampling.

Second, we demonstrate a simple deterministic algorithm that enables one to

efficiently identify the most important matrix elements in a column, without wasting time iterating through all of them, as all determinant-based algorithms had done before [24]. We use this algorithm to efficiently perform Selected CI by finding only the most important Slater determinants to add to a growing determinant space on each iteration. We call this new efficient Selected CI algorithm Heat-bath Configuration Interaction (HCI).

The outline of this thesis is as follows. In Chapter 2, I give an overview of quantum chemistry, focusing on the algorithms representing the state-of-the-art prior to the work presented in this thesis. In Chapter 3, I describe Semistochastic Quantum Monte Carlo, the combination of stochastic and deterministic methods of projecting onto the ground state. Chapters 4 and 5 are devoted to the stochastic and deterministic manifestations, respectively, of my new Heat-bath idea. Finally, in Chapter 6, I conclude and describe my planned future research directions.

CHAPTER 2

QUANTUM CHEMISTRY BACKGROUND

In this section we give an introduction to quantum chemistry algorithms, starting from Slater determinants, and going up to some of the state of the art algorithms prior to the advances presented in this thesis. Some of these results will be repeated in the following chapters, which are adapted from published works.

2.1 Quantum Chemistry Hamiltonian

In most molecular situations, the electronic and nuclear degrees of freedom can be decoupled, owing to the vast discrepancy in their masses. Because atomic nuclei are so massive compared to electrons, it is often a good approximation to assume that the nuclei are infinitely-massive point charges, and only the electrons need to be treated quantum mechanically. This is known as the Born-Oppenheimer approximation, and it is used extensively in quantum chemistry.

Also, the electrons in molecules tend to move very slowly compared to the speed of light, the innermost (fastest) electrons moving at a speed of approximately $Nc/137$, where N is the atomic number. Therefore, usually, relativistic effects can be ignored, or at least added in perturbatively, for example by using the Douglas-Kroll Hamiltonian. Another option for approximately including relativistic effects is to take advantage of the fact that that the effect on the valence electrons is primarily an indirect effect due to it changing the distribution of core electrons, so that relativistic effects can be built into the pseudo-potential to a reasonable approximation.

The Hamiltonian we are interested in for molecular systems, therefore, is the

non-relativistic quantum chemical Hamiltonian in the Born-Oppenheimer approximation, i.e.,

$$\hat{H} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j<i}^{N} \frac{1}{r_{ij}},$$

where N and M are the numbers of electrons and nuclei, respectively.

2.2 Second quantization

In attempting to find approximate eigenstates of the above Hamiltonian, it is convenient to introduce a set of orthonormal single-particle orbitals, $\{\phi_i\}$. These single-particle orbitals are usually chosen to be linear combinations of Gaussians centered at the atomic nuclei. Classes of basis sets have been developed for which extrapolation to the complete basis set limit is smooth. For example, the correlation-consistent basis sets of Dunning et al are the most commonly used. They constitute a series of basis sets with increasing valence flexibility (each core orbital is described by only one basis function), and are labeled cc-pVDZ (correlation-consistent polarized valence double zeta), cc-pVTZ (T=triple), etc., where the "zeta" index corresponds to the number of shells the valence electrons can excite from and to. For example, for the carbon atom, the core consists of one s orbital, and the valence shells contain 1s1p, 1s1p1d, etc. For the silicon atom, the core consists of 2 s orbitals and 1 p orbitals, and the valence shells are again 1s1p, 1s1p1d, etc. Therefore, for carbon, the "minimal basis" (single-zeta) would consist of 5 orbitals (3s2p), the cc-pVDZ basis (double-zeta) would consist of 14 orbitals (3s2p1d), the cc-pVTZ basis would consist of 30 orbitals (4s3p2d1f), and so on. For silicon, the minimal basis (single-zeta) would consist of 9 orbitals (2s1p), cc-pVDZ would consist of 18 orbitals (4s3p1d), cc-pVTZ would consist of 34 orbitals (5s4p2d1f), etc. These basis sets were developed to enable smooth extrapolation

to the infinite-basis set limit, and are used extensively in this work.

Rotations between the orbitals preserve the space that they span. Therefore, both the basis set and the particular choice of orbital rotations must be specified, unless the exact solution within the basis is obtained (called Full CI, to be described later).

2.2.1 Slater determinants

Once the single-particle basis and orbital rotations are chosen, we can then construct multi-particle states as products of these single-particle states. However, because the many-electron wavefunction is anti-symmetric under interchange of a pair of electrons, we only consider linear combinations of these product states that contain this anti-symmetry. Mathematically, this is represented as a determinant, referred to as a Slater determinant,

$$D\left(r_1, r_2, ...r_N\right) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_{i_1}\left(r_1\right) & \phi_{i_1}\left(r_2\right) & \cdots & \phi_{i_1}\left(r_N\right) \\ \phi_{i_2}\left(r_1\right) & \phi_{i_2}\left(r_2\right) & & \\ \vdots & & \ddots & \\ \phi_{i_N}\left(r_1\right) & & & \phi_{i_N}\left(r_N\right) \end{vmatrix}, \tag{2.1}$$

where $\left\{\phi_{i_j}\right\}$ represent the basis functions that are occupied in the current Slater determinant.

2.2.2 Second-quantized Hamiltonian

In this basis of Slater determinants, the quantum chemical Hamiltonian takes on a very simple form:

$$\hat{H} = \sum_{pr} f_{rp} a_r^\dagger a_p + \frac{1}{2} \sum_{pqrs} g_{rspq} a_r^\dagger a_s^\dagger a_q a_p + h_{\text{nuc}}, \tag{2.2}$$

where $a_p^\dagger(a_p)$ is the usual electron creation (annihilation) operator with the indices $\{p, q, r, s\}$ incorporating both spatial and spin degrees of freedom. The tensors entering the expression of the Hamiltonian are the one-electron (two-center) integrals,

$$f_{rp} = \int \phi_r^*(x) \left(-\frac{1}{2}\nabla^2 - \sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{r}_I|} \right) \phi_p(x)\, dx, \tag{2.3}$$

with $\phi_p(x)$ denoting spin-orbitals, x the combined spatial ($\mathbf{r}$) and spin coordinates of the electrons, and Z_I and $\mathbf{r}_I$ the atomic number and spatial coordinates of nucleus I; the two-electron (four-center) integrals,

$$g_{rspq} = \int \phi_r^*(x_1)\, \phi_s^*(x_2)\, \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_p(x_1)\, \phi_q(x_2)\, dx_1 dx_2, \tag{2.4}$$

with an index-ordering convention according to the physicist notation [2]; and the nuclear-nuclear repulsion,

$$h_{\text{nuc}} = \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{r}_I - \mathbf{r}_J|}. \tag{2.5}$$

2.3 Hartree-Fock Theory

Even in the basis of Slater determinants, the many-body Schrodinger equation is still difficult to solve, in part due to the fact that every electron interacts with every other electron.

A convenient starting point in finding approximate solutions to the many-body Schrodinger equation is to approximate this electron-electron interaction as an interaction between an electron and the mean electron field created by the other electrons. This enables the challenging N-body problem to be reduced to N one-body problems.

Its ground state is a single Slater determinant. Hartree-Fock theory can alternatively be interpreted as a variational method where the wavefunction ansatz is a single Slater determinant and the orbital rotations are the variational parameters. The mean-field equations for the Hartree-Fock orbitals are the integro-differential equations [38]

$$h(1)\,\phi_a(1) + \sum_{b\neq a}\left[\int d\mathbf{r}_2\,|\phi_b(2)|^2\,\frac{1}{r_{12}}\right]\phi_a(1) + \sum_{b\neq a}\left[\int d\mathbf{r}_2\phi_b^*(2)\,\phi_a(2)\,\frac{1}{r_{12}}\right]\phi_b(1) = \epsilon_a\phi_a(1) \quad (2.6)$$

where $h(1) = -\frac{1}{2}\nabla_1^2 + \sum_A \frac{Z_A}{r_{1A}}$ is the one-body part of the true Hamiltonian, incoporating both an electron's kinetic energy and its Coulomb attraction to atomic nuclei. The above equation yields the orbital energies $\{\epsilon_a\}$ of the orbitals $\{\phi_a\}$.

2.4 Full Configuration Interaction (Full CI)

Of course, the Hartree-Fock determinant is only the simplest variational wavefunction and a convenient starting point for more computationally-intensive methods. The full solution of the quantum chemical Hamiltonian within a given basis of Slater determinants is called Full Configuration Interaction. In the limit of a complete single-particle basis set, the Full CI solution converges to the exact ground-state wavefunction.

Full CI just corresponds to finding the lowest eigenvector of the Hamiltonian matrix in the basis of Slater determinants. However, the number of Slater determinants scales exponentially with the system size (both the number of electrons and the number of orbitals),

$$N_{FCI} \sim \binom{M}{n_\uparrow}\binom{M}{N-n_\uparrow},$$

where M is the number of spatial orbitals, N is the number of eletrons, and $n_\uparrow$ is the number of spin-$\uparrow$ electrons. Therefore, an exact, deterministic solution is not possible except for the smallest systems.

2.4.1 Deterministic methods

One way of obtaining the Full CI energy is by exact diagonalization of the Hamiltonian, which is an $\mathcal{O}\left(N_{FCI}^3\right)$ operation. However, it should be noted that this is overkill; exact diagonalization gives the entire spectrum of eigenstates of the Hamiltonian, whereas we are only interested in the ground state (and maybe a few low-lying excited states). Therefore, we here consider iterative deterministic methods that aim to obtain only the ground state, or only the lowest few states, which is far simpler than obtaining the entire eigenspectrum.

Power method

Conceptually the simplest iterative method is based on the fact that the eigenstate of interest is an extremal one. Consider a matrix P and an initial starting vector u. If one applies the matrix P repeatedly to the starting vector u, the dominant eigenvector (the one corresponding to the largest eigenvalue of P) will be projected

out:

$$\begin{aligned}\lim_{n\to\infty} P^n u &= \lim_{n\to\infty} P^n \sum_i c_i v_i \\ &= \lim_{n\to\infty} \sum_i \lambda_i^n c_i v_i \\ &\propto v_{\max},\end{aligned}$$

where $\{\lambda_i\}$ and $\{v_i\}$ are the eigenvalues and eigenvectors of P, respectively, and $v_{\max}$ is the eigenvector corresponding to the maximum eigenvector of P. On each application of P, all the components of u perpendicular to $v_{\max}$ decrease in magnitude with respect to the component along $v_{\max}$.

Of course, we are interested in the smallest eigenvalue of the Hamiltonian, not the largest. Therefore, we choose P, which we dub the projector, to be an operator that inverts the spectrum of H. The simplest choice is the linear projector,

$$P_{ij} = \delta_{ij} - \tau \left(H_{ij} - E_T \delta_{ij}\right),$$

where $\tau < \frac{2}{E_{\max} - E_{\min}}$ and E_T is chosen to be close to the true ground state E_0 so that the eigenvalue of the projector for that state is close to 1.

The power method does not require one to store H; it only requires the storage of two u vectors, representing the approximation to the ground state wavefunction for the current and most recent iterations. However, it can take many iterations to fully converge.

Lanczos method

Instead of a priori choosing the projector at iteration n to be given by $P_n = P^n = (1 - \tau (H - E))^n$, a smarter choice is to declare that the projector be an nth-order

polynomial in H with coefficients to be optimized, i.e.,

$$P_n = \sum_{i=0}^{n} k_i H^i.$$

Then, one can optimize the coefficients $\{k_i\}$ so that the projected state $P_n u$ has minimal energy. This is known as the Lanczos method.

In practice, this optimization can be performed as a simple linear algebra problem. First, define the so-called Krylov space spanned by

$$\left\{u, Hu, H^2u, ...H^nu\right\}.$$

We are interested in the lowest-energy state within this space. For simplicity, we first perform the Gram-Schmidt procedure to orthonormalize the Krylov vectors with respect to each other. Then, compute the $n \times n$ Hamiltonian within the orthonormalized Krylov space, and find its lowest eigenvalue. This is the energy of the lowest-energy state that can be projected out of initial vector u using an nth-order projector P_n.

The same procedure can be repeated at increasing values of n, until convergence is achieved. This convergence comes much faster than in the case of the power method because of the increased flexibility in the form of the projector P_n.

2.4.2 Stochastic method: Full CI Quantum Monte Carlo

The stochastic implementation of the power method, called Full CI Quantum Monte Carlo (FCIQMC), will be described in detail in Chapter 3. Here we just give a brief introduction. The deterministic Full CI methods all require one to be able to store a vector of the size of the number of Slater determinants, so they can only be used on small systems. One can get around this storage requirement by

introducing stochasticity: instead of storing the full vector each iteration, one only needs to store a sparse "snapshot" that is equal to the full vector in expected value. The sparsity of these "snapshots" can be maintained by stochastically simulating the matrix-vector multiplication of the power method, by randomly zeroing out most of the off-diagonal elements of the Hamiltonian each iteration, then scaling the nonzero elements appropriately to maintain the correct expected value. By updating an estimate of the ground state energy each iteration, only a single sparse "snapshot" needs to be kept around at a time, and statistics can be accumulated over the course of a long run including many "snapshots" to obtain an accurate estimate of the energy and its uncertainty.

2.5 Correlation Energy

We now take a break from our discussion of algorithms to discuss the so-called correlation energy, the energy resulting from the fact that the single-particle electronic states are correlated with one another, so that the N-particle wavefunction cannot be factored into N single-particle states. This energy is defined as the difference between the Hartree-Fock energy and the Full CI energy. By approximating the electronic Hamiltonian in a mean-field way, Hartree-Fock misses out on the energy of electronic correlation in two ways, often referred to as dynamic and static correlation.

2.5.1 Dynamic Correlation

Dynamic correlation is due to the correlated motion of the electrons. By approximating the interaction between all pairs of electrons as the interaction between each electron and an average of all the other ones, there is no Coulomb interaction preventing pairs of electrons from getting too close to each other. Therefore, the Hartree-Fock solution will have an energy too high because it will have an inflated probability of two electrons being close, where they will have a large repulsion energy. Furthermore, the wavefunction should have a cusp where the electrons coincide, which is entirely absent in the Hartree-Fock wavefunction.

Dynamic correlation is usually thought of as a sum of contributions of many small terms, representing excitations into the unoccupied orbitals. It is often added in after the fact to a given reference wavefunction (which may be a single determinant or a relatively small linear combination of determinants), e.g., with perturbation theory or with an excitation operator as in coupled cluster theory.

2.5.2 Static Correlation

Static correlation, on the other hand, comes from the fact that for some systems the Hartree-Fock ground state has the wrong qualitative features relative to the true Full CI ground state. The ground state of the Hartree-Fock Hamiltonian is a single Slater determinant, namely the Hartree-Fock determinant, whereas, for some systems, e.g., for molecules at a stretched geometry, several determinants may have sizable coefficients in the exact wavefunction.

Unlike dynamic correlation, important contributions to the static correlation

generally come from a reltively small number of Slater determinants. Hence, it is often possible to get an accurate description of static correlation by using a multi-determinant wavefunction. How one might identify the determinants to add to the Hartree-Fock determinant is the subject of the next section.

2.6 Truncated CI

Since the full set of Slater determinants is too large to store for all but the smallest systems, we here consider the ground state within a truncated subset of the full determinant space. This selected CI solution will not only provide us with a variational upper bound to the Full CI energy, but can also serve as a multi-reference wavefunction incorporating more static correlation than the Hartree-Fock determinant.

The basic idea is to identify a set of Slater determinants, compute the Hamiltonian between pairs of them, then find the lowest eigenvalue and corresponding eigenvector of this truncated Hamiltonian. There are three primary ways of identifying new Slater determinants:

1. Limit the number of excitations from the Hartree-Fock determinant. For example, consider all determinants that are either single or double excitations from the HF determinant, and diagonalize in this space. This is known as CI Singles and Doubles (CISD).
2. Divide the set of orbitals into three categories: Core (always occupied), Active (sometimes occupied), and Virtual (never occupied). Then, consider all determinants that meet these criteria. This is known as a Complete Active

Space (CAS) calculation. One can also simultaneously optimize rotations between orbitals in different categories to further increase the variationality, yielding what is known as a CAS Self-Consistent Field (CASSCF). Assuming the set of active orbitals is chosen well, a CAS wavefunction is great at describing static correlation, since it includes most of the important determinants (determinants with excitations out of the core or into the virtual space are less important for qualitatively describing the ground state). However, a CAS or CASSCF wavefunction is not good at describing dynamic correlation, since a sum over the many small contributions from excitations into the virtual space is needed to correctly describe the repulsion between electrons.

3. Iteratively add new determinants to a growing list of important ones. This is known as Selected CI and will be treated in a later section.

2.7 Perturbation Theory

The idea of perturbation theory is to find the lowest-order correction to a reference wavefunction. We write the Full CI Hamiltonian H as a sum of two pieces, a zeroth-order Hamiltonian H_0, of which the reference wavefunction constitutes the ground state, and a "perturbation," V:

$$H = H_0 + V.$$

Letting $\psi^{(0)}$ and $E^{(0)}$ denote the reference wavefunction and energy, the lowest-order perturbative correction to the energy occurs at second order, and is given by

$$E^{(2)} = \left\langle \psi^{(0)} \middle| V \frac{1}{E^{(0)} - H_0} V \middle| \psi^{(0)} \right\rangle.$$

This formula is valid for any choice of partitioning of the Hamiltonian into its zeroth-order and perturbative components. Here we consider a few enticing choices.

2.7.1 Epstein-Nesbet Perturbation Theory

Conceptually the simplest choice of partitioning is for the zeroth-order Hamiltonian to include the block connecting pairs of determinants in the reference, as well as all other diagonal matrix elements. In that case, the second-order perturbative correction simplifies to

$$\begin{aligned} E^{(2)} &= \left\langle \psi^{(0)} \middle| V \frac{1}{E^{(0)} - H_0} V \middle| \psi^{(0)} \right\rangle \\ &= \sum_k \frac{\left| \left\langle k \left| V \right| \psi^{(0)} \right\rangle \right|^2}{E^{(0)} - H_{kk}}. \end{aligned}$$

Epstein-Nesbet perturbation theory is fast because the part of the perturbation that appears in the denominator (namely, $E^{(0)} - H_0$), is diagonal, so that inversion of that operator is trivial. However, diagonal elements of the Hamiltonian are expensive to compute. Computing a diagonal matrix element from scratch requires $\mathcal{O}\left(N_{\text{el}}^2\right)$ time, since it requires summing over all pairs of electron-electron interactions. There is no reason to ever compute a diagonal matrix element for the perturbation theory expression from scratch, though, since each determinant whose diagonal element we need is a single or double excitation from a determinant whose diagonal element we already know, and the difference between the two diagonal matrix elements can be computed in only $\mathcal{O}\left(N_{\text{el}}\right)$ time by summing over only the pairs that changed in the interaction (all the ways that the one or two exctited electrons can be paired with other electrons that did not excite).

2.7.2 Moller-Plesset Perturbation Theory

While Epstein-Nesbet perturbation theory is fast, it still takes $\mathcal{O}(N_{\text{el}})$ time to compute each determinant that contributes to the perturbation theory expression. Moller-Plesset perturbation theory aims to reduce this computational cost. For single-reference Moller-Plesset perturbation theory, the diagonal component of the perturbation V from Epstein-Nesbet perturbation theory, namely H_{kk}, is replaced by an approximate expression that can be computed in $\mathcal{O}(1)$ time. This is done by approximating the change in diagonal matrix element from the reference determinant to an excited determinant as just the change in the orbital energies of the orbitals that changed occupancy in the excitation. The zeroth-order Hamiltonian is then modified appropriately, so that there are in reality no approximations made in the partitioning of the Hamiltonian. Thus, in Moller-Plesset perturbation theory, H_0 is diagonal, with elements given by

$$(H_0)_{ii} = H_{11} + \Delta\epsilon_{\text{orb}}$$

where $\Delta\epsilon_{\text{orb}}$ is the difference in the orbital energies from state 1 to state i. In this case, the second-order Moller-Plesset perturbation theory (MP2) expression is given by

$$E_{\text{MP2}} = \sum_{p,q\in\text{occ.}} \sum_{r,s\in\text{unocc.}} \frac{|\langle pq|rs\rangle|^2}{\epsilon_p + \epsilon_q - \epsilon_r - \epsilon_s},$$

where we have made use of the fact that there are no single excitations from the Hartree-Fock reference.

2.7.3 Linearized Coupled Cluster Theory

When the reference wavefunction is a CAS or CASSCF wavefunction, there is a particularly interesting way to partition of the Hamiltonian [35, 34]. When

constructing a CAS, the orbitals are divided into three categories: core, active, and virtual. The zeroth-order Hamiltonian can then be chosen to be only the terms in the Hamiltonian for which the total number of electrons in each category remain unchanged. This means that the zeroth-order Hamiltonian is block-diagonal, and the CAS wavefunction is the ground state within the block for which the core orbitals are all occupied and the virtual ones are all unoccupied.

One of the advantages of this partitioning is that the zeroth-order Hamiltonian is close to the true Hamiltonian, so the second-order energy tends to be a better estimate of the true energy than in other forms of perturbation theory. However, it should be noted that this comes with a price: the zeroth-order Hamiltonian is no longer easy to invert, since it is not diagonal.

2.8 CI by Perturbatively Selecting Iteratively (CIPSI)

Since truncated CI is useful for describing static correlation, while perturbation theory is useful for describing dynamic correlation, it is useful to consider combining the two in order to address both types of electron correlation. Of course, one way to do this is by choosing the truncated CI part to be, for example, a CASSCF wavefunction. However, a more flexible approach is to iteratively add new determinants to a growing list of important ones, as in Selected CI. There are now many algorithms that combine a Selected CI algorithm with Perturbation Theory. Here we review the first such algorithm, known as Configuration Interaction by Perturbatively Selecting Iteratively (CIPSI) [25].

CIPSI consists of two stages: the variational stage, in which a variational wavefunction is generated, and the perturbative stage, in which a perturbative correc-

tion to the variational wavefunction is computed.

2.8.1 Variational stage

The variational wavefunction is the lowest-energy eigenstate of the Hamiltonian within a selected space of determinants. The selected space is chosen using an iterative process as follows.

The initial selected space of determinants consists of only the HF determinant. Then, each iteration:

1. The current variational wavefunction is obtained as the lowest-energy eigenstate of the Hamiltonian within the current selected space.
2. For each determinant connected to the current variational wavefunction, first-order perturbation theory is used to estimate of the coefficient the determinant would have if added to the variational wavefunction.
3. Then, those determinants with largest-magnitude expected coefficients are added to the variational space.

The process is repeated until either convergence or computational limits are reached.

2.8.2 Perturbative stage

Once a variational wavefunction has been obtained as a linear combination of determinants, Epstein-Nesbet perturbation theory is used to compute the lowest-

order (second-order) correction to the variational energy. Of course, while the variational energy is an upper bound on the true Full CI energy, there is nothing preventing the perturbative correction from "overshooting," and it can be either above or below the Full CI energy.

2.8.3 Inefficiencies

It should be noted that although CIPSI has long been a widely-used algorithm, it has some very obvious inefficiencies. First, in the variational stage, *all* determinants connected to the variational wavefunction by nonzero Hamiltonian matrix elements are iterated over, even though the vast majority will not be added to the selected determinant space. Also, in the perturbative stage, *all* connected determinants are again iterated over, since they all contribute to the perturbative correction, even though many contribute a negligible amount.

My new algorithm, Heat-bath Configuration Interaction (HCI), solves both of these problems, to dramatically increase the efficiency. HCI will be the subject of Chapter 5 of this thesis.

CHAPTER 3

SEMISTOCHASTIC QUANTUM MONTE CARLO

Here I introduce a *semistochastic* implementation of the power method for projecting out extremal eigenvalues of large matrices. At different times throughout this thesis, I refer to this algorithm as either Semistochastic Full Configuration Interaction Quantum Monte Carlo (S-FCIQMC) or Semistochastic Quantum Monte Carlo (SQMC). The method is semistochastic in the sense that the matrix-vector multiplication of the power method is carried out partly deterministically and partly stochastically. The motivation for this semistochastic algorithm comes from the fact that both deterministic and stochastic projection algorithms have their respective advantages and shortcomings. Deterministic algorithms, while exact, are limited to very small Hilbert spaces of determinants (about 10^{10}), while stochastic algorithms get around this limitation at the expense of introducing a statistical uncertainty and bias. Semistochastic Quantum Monte Carlo combines the best of both worlds: it can be used on problems as large as stochastic algorithms can, while at the same time dramatically reducing the statistical uncertainty and bias by performing the most important part of the projection deterministically. This chapter was partially adapted from our recent paper published in Physical Review Letters [31].

3.1 Introduction

We are interested in the general problem of finding the ground state energy of a quantum system in a discrete Hilbert space of N Slater determinants. This amounts to solving the time-independent Schrodinger equation,

$$\hat{H}\Psi_n = E_n\Psi_n,$$

where $\hat{H}$ is the $N \times N$ (Hermitian) Hamiltonian matrix in the space of Slater determinants, $\{E_n\}$ are the energy eigenvalues, and $\{\Psi_n\}$ are the corresponding stationary states. We assume that $E_0 < E_1 \leq \cdots \leq E_{N-1}$, i.e., that there is a non-degenerate ground state energy E_0. Computing the ground state Ψ_0 in this manner, by finding the lowest-energy linear combination of the *full* set of Slater determinants constructed from the given basis, is referred to as Full Configuration-Interaction (Full CI) in theoretical quantum chemistry.

For a small enough Hilbert space, this ground state energy can be calculated deterministically, but we want a method that will work for a Hamiltonian in any Hilbert space which can be arbitrarily large - even Hilbert spaces that far exceed memory limitations. In this chapter, we investigate, for completeness, both deterministic and stochastic methods for estimating the ground state energy, both of which have advantages and disadvantages. We then explore a semistochastic projection algorithm that we invented, which combines the best qualities of both deterministic and stochastic projection. This method, which we have dubbed Semistochastic Quantum Monte Carlo (SQMC), as well as some improvements and applications, will be the focus of the rest of this chapter.

3.2 Deterministic Projection

In a Hilbert space small enough that the entire Hamiltonian can be stored, exact diagonalization is trivial. In a Hilbert space large enough that this is impossible, deterministic methods such as the power method and the Lanczos algorithm can be used, provided that we can still store two vectors of length N.

3.2.1 Power method

Since we are only interested in the ground state, or lowest eigenvalue of $\hat{H}$, we don't have to solve the full eigenvalue problem, and we can instead consider algorithms that just calculate one of the extremal eigenvalues. One such algorithm is the power method, which projects out the dominant eigenvector (i.e., the eigenvector corresponding to the largest-magnitude eigenvalue) of a matrix $\hat{P}$ by repeatedly multiplying $\hat{P}$ by an arbitrary starting vector $v^{(0)}$,

$$\begin{aligned}
\lim_{n\to\infty} \hat{P}^n v^{(0)} &= \lim_{n\to\infty} \hat{P}^n \sum_{k=1}^{N} v_k \left(v_k \cdot v^{(0)}\right) \\
&= \lim_{n\to\infty} \sum_{k=1}^{N} \lambda_k^n v_k \left(v_k \cdot v^{(0)}\right) \\
&\propto v_{\mathrm{dom}},
\end{aligned}$$

where we have assumed that the eigenvectors are orthonormal and that the starting vector $v^{(0)}$ has nonzero overlap with the dominant eigenvector v_{dom}. We call $\hat{P}$ the "projector," even though in reality, only $\hat{P}_\infty = \lim_{n\to\infty} \hat{P}^n$ actually projects onto the dominant eigenvector. Since we are interested in the lowest eigenvalue of the Hamiltonian $\hat{H}$, we choose a projector $\hat{P}$ such that its dominant eigenvector corresponds to the lowest eigenvector of $\hat{H}$.

3.2.2 Lanczos algorithm

The Lanczos algorithm is a method similar to the power method, but it converges much faster. It can also approximate both the highest and lowest eigenvalues simulataneously.

Consider a projector as the first L terms of a Taylor series in $\hat{H}$ with undetermined coefficients $\{c_i\}$,

$$\hat{P}_L^{\{c_i\}} = \sum_{i=0}^{L-1} c_i \hat{H}^i.$$

The Lanczos algorithm minimizes the energy of

$$v_L^{\{c_i\}} = \hat{P}_L^{\{c_i\}} v^{(0)}$$

for some initial starting vector v_{init}:

$$E_L = \min_{\{c_i\}} \frac{\left\langle v_L^{\{c_i\}} \middle| \hat{H} \middle| v_L^{\{c_i\}} \right\rangle}{\left\langle v_L^{\{c_i\}} | v_L^{\{c_i\}} \right\rangle}.$$

Alternatively, E_L can be computed as the lowest eigenvalue of the Hamiltonian in the basis $\left\{\hat{H}^i v^{(0)}\right\}_{i=0}^{L-1}$. This $L \times L$ Hamiltonian is given by (suppressing the superscript of $v^{(0)}$ for brevity)

$$H_L = \begin{pmatrix} \langle v | H | v \rangle & \langle v | H^2 | v \rangle & \cdots & \left\langle v \middle| H^L \middle| v \right\rangle \\ \langle v | H^2 | v \rangle & \langle v | H^3 | v \rangle & & \\ \vdots & & \ddots & \\ \left\langle v \middle| H^L \middle| v \right\rangle & & & \left\langle v \middle| H^{2L-1} \middle| v \right\rangle \end{pmatrix}.$$

The algorithm can be made even more efficient by first performing a Gram-Schmidt procedure on the vectors $\left\{\hat{H}^i v_{\text{init}}\right\}_{i=0}^{L-1}$, so that H_L becomes a tri-diagonal matrix.

In the limit $L \to N$, the projector $\hat{P}_L^{\{c_i\}}$ has all the flexibility necessary to exactly project v_{init} onto the true ground state v_0. However, in practice, E_L typically gets to within machine precision of E_0 in $L \sim 10^2$ even for $N \sim 10^5$.

Note that this algorithm can also be used to compute the largest possible eigenvalue, by taking the largest eigenvalue of H_L. It can also be used to estimate the first excited state, by taking the second-lowest eigenvalue. This would typically require a much larger value of L to converge, however.

3.3 Stochastic Projection

For the case where N is too large to store even a single vector of that length, we can no longer perform the projection deterministically and must resort to a stochastic method. Note that stochastic projection really is a last resort. If N is small enough that we have a choice between deterministic and stochastic projection methods, we would always choose deterministic, which gives us the exact answer without any statistical uncertainty. Stochastic projection, by contrast, estimates the ground state energy by sampling the ground state wavefunction, which not only has statistical uncertainty, but also runs into the infamous Fermion sign problem, as we will see.

The method described below is referred to as Full Configuration-Interaction Quantum Monte Carlo (FCIQMC), since it is used to compute the Full CI energy in a given basis. It is a stochastic implementation of the power method.

The power method in its original (deterministic) form cannot be used here because we would have to store the vector

$$v^{(t)} = \hat{P} v^{(t-1)}$$

for each iteration of applying the projector, and N is too large to do so. Therefore, we seek an algorithm that only ever requires us to store a *sparse* representation of

vector $v^{(t)}$.

Suppose we start the run with a sparse initial vector $v^{(0)}$, which has one nonzero element, the Hartree-Fock (HF) determinant, with a value of 1. If we deterministically compute the next iteration's vector $v^{(1)} = \hat{P}v^{(0)}$, we find that it is much less sparse. In fact, $v^{(1)}$ is equal to the column of $\hat{P}$ corresponding to the HF determinant. The Hamiltonian matrices we deal with tend to be sparse, but far from diagonal: the number of nonzero elements per column is typically $N_{off} \sim 10^2 - 10^4$. Thus, if we applied the projector deterministically, even though we could probably store $v^{(1)}$, after just a few iterations, we would already have more nonzero elements than we could store.

We therefore *simulate* the matrix-vector multiplication $\hat{P}v^{(0)}$ stochastically such that its expectation value is $v^{(1)}$, in a way that preserves sparsity. The most naive way to accomplish this is to allow projection via only one of the nonzero off-diagonal elements - selected uniformly - and then multiply the resulting amplitude by the number of such nonzero off-diagonal elements. The diagonal projector element can be applied deterministically, since it already has nonzero value and thus doesn't lose much sparsity. It is trivial to verify that the expectation value of this stochastic matrix-vector multiplication algorithm is the same as the result of a deterministic matrix-vector multiplication. The resulting vector is nearly as sparse as the starting vector: the number of nonzero elements has only been increased by a factor of 2, rather than by a factor of $(N_{off} + 1) \gg 2$. This naive algorithm for stochastic matrix-vector multiplication is illustrated in Figure 3.1.

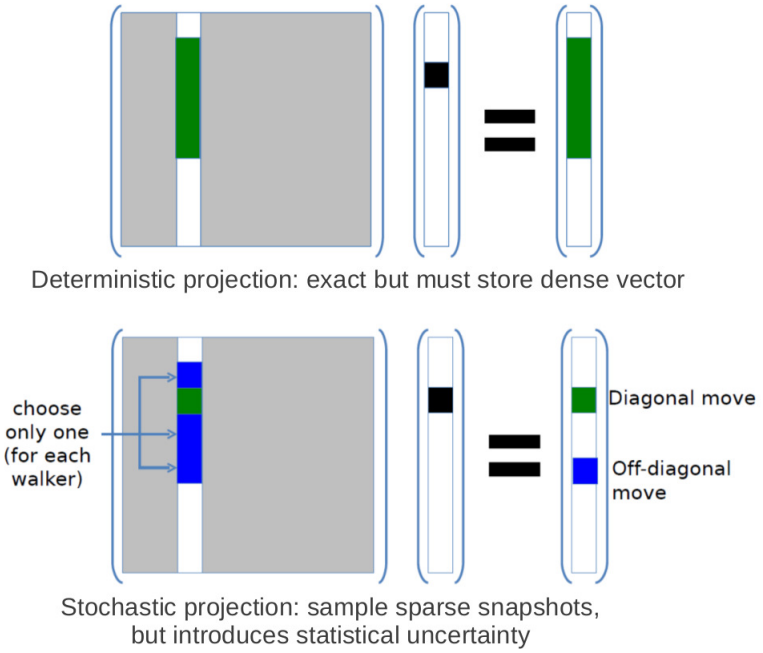

Figure 3.1: Diagram showing pictorially the matrix-vector multiplication performed deterministically and stochastically. Green represents deterministic parts and blue represents stochastic parts of the projection. White represents elements that are zero, and the vector that is being multiplied has only a single nonzero element, indicated in black. When multiplying such a sparse vector by a matrix, only the corresponding column of the matrix is needed, so the rest of the matrix is grayed out. After a deterministic matrix-vector multiplication, the resulting vector is much less sparse than the original one. In stochastic matrix-vector multiplication, the diagonal element is applied deterministically, but only one off-diagonal element is selected randomly to be applied, and its value divided by the probability of selecting it. The resulting vector has expectation value equal to the vector that results from deterministic matrix-vector multiplication, without losing much sparsity.

3.3.1 Walkers

A less noisy way to perform stochastic projection is to divide up the weight into an integer number of "walkers," and have each walker independently make an off-diagonal move. We then can represent the sparse vector at any iteration as a population of signed walkers, each of which occupies a Slater determinant:

$$v^{(t)} = \sum_i w_i^{(t)} |i\rangle ,$$

where $w_i^{(t)}$ is the weight on determinant $|i\rangle$ on iteration t. Note that this sparse encoding only requires storage of the labels of nonzero-weight (currently occupied) determinants and their current weights on any given iteration.

At this point, since we are deriving the dynamic that walkers must follow, it is helpful to go ahead and choose a specific projector to use. Recall that in the power method, we need to choose a projector $\hat{P}$ such that its dominant eigenvector corresponds to the lowest eigenvector of $\hat{H}$. The simplest such choice is the linear projector,

$$\hat{P}_{ij} = \delta_{ij} - \tau \left(\hat{H}_{ij} - E_{\mathrm{T}} \delta_{ij} \right) ,$$

where E_{T} is chosen to be close to E_0 so that the weight on the ground state stays about the same, and the time step τ must be chosen such that the smallest eigenvalue is lower in magnitude than the largest one, i.e., $\tau < \frac{2}{\text{spectrum}}$. We will use this simplest choice for much of the thesis.

Using the linear projector, the matrix-vector multiplication that we would like

to simulate stochastically is given by

$$\begin{aligned} w_i^{(t)} &= \sum_j \hat{P}_{ij} w_j^{(t-1)} \\ &= \sum_j \left[\delta_{ij} - \tau\left(\hat{H}_{ij} - E_{\mathrm{T}}\delta_{ij}\right)\right] w_j^{(t-1)} \\ &= \left(1 - \tau\left(\hat{H}_{ii} - E_{\mathrm{T}}\right)\right) w_i^{(t-1)} - \tau \sum_{j \neq i} \hat{H}_{ij} w_j^{(t-1)}. \end{aligned}$$

Note that the first of these terms is diagonal, and the second is off-diagonal. Recall that in our stochastic implementation of the power method, the diagonal elements can be applied deterministically, since they do not decrease the sparsity of our sparse representation. It is the off-diagonal terms which we would like to simulate stochastically.

We therefore use the following walker dynamic. For each occupied determinant:

1. Divide up the weight evenly into an integer number of walkers whose absolute weights are approximately one.
2. For each walker i with walker weight $w_i^{(t-1)}$ ($\approx \pm 1$), choose a single determinant j to propose an off-diagonal move to, with probability $P(j \leftarrow i)$.
3. Spawn a new walker on determinant j with weight

   $$w_j^{(t)} = \frac{-\tau H_{ji} w_i^{(t-1)}}{P(j \leftarrow i)}.$$

   This is the stochastic off-diagonal move.
4. Multiply the weights that were there already there at the end of the last iteration by their diagonal elements:

   $$w_i^{(t)} = \left(1 - \tau\left(\hat{H}_{ii} - E_{\mathrm{T}}\right)\right) w_i^{(t-1)}.$$

5. Merge the newly spawned walkers with the old walker list.

6. Integerize small-weight determinants: If a determinant has an absolute weight $|w_i|$ that is less than the minimum weight allowed w_{min} (usually chosen to be 0.5), then with probability $|w_i|/w_{min}$, scale up its absolute weight to w_{min}; else, discard the determinant.

The last step is necessary to maintain sparsity by removing small-weight determinants.

3.3.2 Proposal probability

We are free to choose the discrete probability distribution that we use to propose off-diagonal moves, since the weight that each walker picks up is divided by the probability of proposing a move there. The ideal distribution is to choose an off-diagonal element with probability proportional to its magnitude, so that the new weight that is spawned is the same no matter what off-diagonal element is selected. However, this requires summing over all the nonzero off-diagonal elements, so we try to approximate the distribution in a more efficient way. This approximate heat-bath sampling algorithm will be the focus of Chapter 4 of this thesis, but or now, we just use the conceptually simplest (but highly inefficient!) choice of sampling distribution, namely, uniform sampling.

3.3.3 Energy estimators

Now that we have a walker dynamic that samples the ground state wavefunction, we still need a method of extracting a ground state energy estimate from it. If we could store the exact wavefunction, we would just use the so-called "pure" energy

estimator,

$$E_0^{\text{pure}} = \frac{\left\langle \psi_0 \middle| \hat{H} \middle| \psi_0 \right\rangle}{\langle \psi_0 | \psi_0 \rangle}.$$

However, we can never store the entire wavefunction ψ_0, but instead only sparse snapshots $v^{(t)}$ such that

$$\psi_0^{\text{MC}} = \frac{1}{N_{\text{MC}}} \sum_t v^{(t)}.$$

In fact, since we will be doing many Monte Carlo iterations, we can't store the entire set of all snapshots, but only one snapshot at a time. Therefore, we need to use an estimator of E_0 that depends only on linear functions of the snapshots. This rules out the pure energy estimator above, since both the numerator and denominator are quadratic functions of the snapshots; each would require the entire list of snapshots to compute. Instead, in our code we use the so-called "mixed" estimator,

$$E_0^{\text{mixed}} = \frac{\left\langle \psi_0^{\text{MC}} \middle| \hat{H} \middle| \psi_{\text{T}} \right\rangle}{\langle \psi_0^{\text{MC}} | \psi_{\text{T}} \rangle},$$

for some trial wavefunction ψ_{T} that we choose before the run. In this mixed energy estimator, we can compute its numerator and denominator separately, since each is a linear function of the snapshots. The algorithm is correct for any ψ_{T} that is not orthogonal to ψ_0, but the fluctuations in the numerator and denominator are much smaller the closer ψ_{T} is to ψ_0. In fact, if ψ_{T} were equal to the true ground state, then the fluctuations in the mixed energy estimator would be zero! In the original formulation of FCIQMC, ψ_{T} was chosen to be the Hartree-Fock state. However, one can easily find a linear combination of a small number of determinants (e.g., CISD) that is much lower in energy and/or closer to ψ_0. We will discuss this and other improvements in section 3.

Another estimator that one could use is the "growth" estimator. Since the eigenvalue of the linear projector corresponding to the ground state is

$1-\tau\left(E_{\mathrm{T}}-E_0\right)$, and since this eigenvalue is equal to 1 when the population is constant, we can use these facts to estimate the ground state energy E_0 as the value of E_{T} that keeps the population constant under the linear projector ($E_{\mathrm{T}} \approx E_0$). One of the drawbacks of this method is that it is not systematically improvable in the way that the mixed energy estimator is. Therefore, we choose to use the mixed energy estimator throughout the rest of this thesis, with a trial wavefunction ψ_{T} optimized prior to the run as described in section 3.2.

3.3.4 Sign problem and initiator approximation

The above formulation of FCIQMC looks good on paper, but unfortunately the infamous "sign problem" rears its ugly head. The sign problem comes from the fact that both $+\psi_0$ and $-\psi_0$ are equally good solutions to the Schrodinger equation. When the power method is applied deterministically, only one of these two solutions is projected out. However, in stochastic projection, without a method of causing the walkers to prefer one over the other, they will sample a linear combination of the two solutions. This leads to an exponentially large noise-to-signal ratio.

One way to get around this, which was proposed by the Alavi group, is the so-called initiator approximation [18]. The idea comes from the fact that the sign problem is caused when wrong-signed determinants propagate, because they will propagate wrong-signed weight. We therefore aim to suppress spawning by wrong-signed determinants. Given that the larger the weight on a determinant, the more confident we are that its sign is correct, we therefore choose to only allow determinants to spawn onto states that are not already occupied if their absolute weight is above a certain threshold. Such determinants are deemed *initiators*.

Under the initiator approximation, the walker dynamic changes as follows. For each occupied determinant:

1. Divide up the weight evenly into an integer number of walkers whose absolute weights are approximately one.
2. For each walker i with walker weight $w_i^{(t-1)}$ ($\approx \pm 1$), choose a single determinant j to propose an off-diagonal move to, with probability $P(j \leftarrow i)$.
3. Spawn a new walker on determinant j with weight

$$w_j^{(t)} = \frac{-\tau H_{ji} w_i^{(t-1)}}{P(j \leftarrow i)}.$$

   This is the stochastic off-diagonal move.

4. Multiply the weights that were there already there at the end of the last iteration by their diagonal elements:

$$w_i^{(t)} = \left(1 - \tau\left(\hat{H}_{ii} - E_{\mathrm{T}}\right)\right) w_i^{(t-1)}.$$

5. Merge the newly spawned walkers with the old walker list. If any determinants contain only weight spawned by non-initiators, discard that weight.
6. Integerize small-weight determinants: If a determinant has an absolute weight $|w_i|$ that is less than the minimum weight allowed w_{min} (usually chosen to be 0.5), then with probability $|w_i| / w_{min}$, scale up its absolute weight to w_{min}; else, discard the determinant.
7. All determinants whose absolute weight is above the initiator threshold are deemed initiators for the next iteration.

The initiator approximation of course biases the walker dynamic and will therefore give us a biased estimate of the ground state energy. However, in the limit of infinite

walker number, all determinants will become initiators, and the initiator bias will disappear. Therefore, when using the initiator approximation, we must run with several walker numbers, and can only trust our ground state energy estimates if the energies appear to be converged with respect to walker number.

3.4 Semistochastic Projection

Deterministic projection methods are exact (no statistical uncertainty or bias), but they can only be used for Hilbert spaces small enough to store. Stochastic projection methods, on the other hand, can be used for arbitrarily large Hilbert spaces, but they come with a statistical uncertainty and an initiator bias for overcoming the sign problem. In this section, we introduce a new method we invented, that combines the advantages of both deterministic and stochastic projection: It can be used for arbitrarily large Hilbert spaces, but the statistical uncertainty and initiator bias are greatly reduced by doing part of the projection deterministically.

3.4.1 Improvement I: Deterministic subspace

Partition the Hilbert space into a two parts, labeled "$\mathcal{D}$," for deterministic, and "$\mathcal{S}$," for stochastic. We will apply the matrix-vector multiplication deterministically between pairs of determinants in $\mathcal{D}$, and otherwise apply it stochastically. We do this by breaking up the projector P into a deterministic piece $P^{\mathcal{D}}$ and a stochastic piece $P^{\mathcal{S}}$:

$$P = P^{\mathcal{D}} + P^{\mathcal{S}},$$

where and $P^{\mathcal{S}} = P - P^{\mathcal{D}}$. Figure 3.2 depicts this division of the projector pictorially.

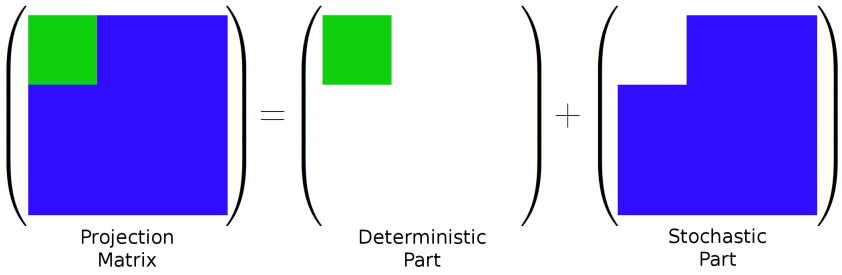

Figure 3.2: Diagram showing pictorially how the projector is broken up into a deterministic piece and a stochastic piece in SQMC. Green represents pieces of the projector that are treated deterministically, and blue represents pieces that are treated stochastically.

The walker dynamic is changed as follows by the inclusion of a deterministic space. The list of determinants within the deterministic space are kept around throughout the whole run; they are not allowed to be discarded even if they have low weight like the other determinants. The deterministic space projector is stored sparsely prior to the run, and on each iteration, new weight equal to $-\tau \sum_{j=1}^{|\mathcal{D}|} (H_{ij} - E_{\mathrm{T}}\delta_{ij})\, w_j$ is deterministically produced on each deterministic determinant. All determinants (including the deterministic ones) are allowed to spawn as usual. However, when the stochastically-spawned new weights are combined with the old weights, any weight which was stochastically spawned between pairs of deterministic determinants is discarded, since this "spawning" was already accounted for deterministically.

3.4.2 Improvement II: Better trial wavefunction

As mentioned before, since we are using a mixed energy estimator which depends on the trial wavefunction ψ_{T}, we can get lower statistical fluctuations if we choose a ψ_{T} that is closer to the true ground state. Whereas FCIQMC was originally formulated

using only Hartree-Fock as the trial wavefunction [12], we created a simple and robust method for generating linear combinations of determinants that are closer to the true ground state [31]. The method for generating such wavefunctions is discussed in the following subsection.

3.4.3 Generating important subspaces

In order to use the above two SQMC improvements, we need to construct both a trial wavefunction ψ_{T} and a deterministic set of determinants $\mathcal{D}$. We choose ψ_{T} to be the ground state eigenvector within a set of determinants $\mathcal{T}$. Thus, we are just in search of a method for generating sets of important determinants, and we can use similar methods for generating both $\mathcal{D}$ and $\mathcal{T}$. There are several methods one could use to generate these two spaces. We chose the following method, which is not only simple, but also flexible enough to allow us to generate spaces of whatever size we want for a particular system and amount of computing power.

An important subspace is generated iteratively as follows.

1. Start with the HF determinant. Generate all connections and find the ground state eigenvector in this space. This is the CISD wavefunction, and this list of determinants becomes the current list for the next step.
2. Take the k determinants from the current list that have the largest absolute weights in the current lowest eigenvector.
3. Generate all connections to these k determinants.
4. Merge this list of connected determinants with the current list of determinants to form a new "current list."

5. Find the lowest eigenvector of the current list of determinants[1].

6. Truncate to the n determinants in this list that have largest absolute weight in the lowest eigenvector.

7. Go to step 2 unless the number of iterations completed (including generating the CISD wavefunction) is already the maximum specified.

The parameters k and n can be different on different iterations. We also have the capability of only generating connected determinants that are excited to a subset of the lowest-lying orbitals, but we rarely use it.

Once the subspace has been generated, if it is used for ψ_{T}, find the lowest eigenvector in the subspace and store that as ψ_{T}. If it is being used for the deterministic subspace, compute the elements of the Hamiltonian between pairs of states in this subspace and store as a sparse matrix.

[1]In practice, we first estimate the coefficients of determinants in the lowest eigenvector using perturbation theory,

$$|\psi_0\rangle \to |\psi_0'\rangle = |\psi_0\rangle + \sum_{k \notin \psi_0} \frac{\langle k|H|\psi_0\rangle}{E_0 - H_{kk}} |k\rangle ,$$

or in our notation,

$$|\psi_0'\rangle = \sum_{k \in |\psi_0\rangle} \frac{N_k}{E_0} |k\rangle + \sum_{k \notin |\psi_0\rangle} \frac{N_k}{E_0 - H_{kk}} |k\rangle .$$

Using these estimates, we first trim the number of determinants down by approximately a factor of 10 to only the determinants whose coefficients in the lowest eigenvector are expected to be the largest. Then, we diagonalize in that space. This procedure reduces the time to generate an important subspace by approximately a factor of 10 without substantially affecting the set of determinants selected.

3.4.4 Saving the Local Energies

Each iteration in the run, we need to update our estimates of the numerator and denominator of the mixed energy

$$\begin{aligned} E_0^{\text{mixed}} &= \frac{\left\langle \psi_0 \left| \hat{H} \right| \psi_{\text{T}} \right\rangle}{\langle \psi_0 | \psi_{\text{T}} \rangle} \\ &= \frac{\sum_{t=1}^{N_{MC}} \sum_i w_i^{(t)} \sum_j H_{ij} t_j}{\sum_{t=1}^{N_{MC}} \sum_i w_i^{(t)} t_i,} \end{aligned}$$

where the last line assumes that we are using a trial wavefunction that is a stored linear combination of Slater determinants,

$$|\psi_{\text{T}}\rangle = \sum_{i \in \mathcal{T}} t_i \, |\phi_i\rangle \, .$$

(For alternative forms of ψ_{T}, see the following section.)

In order to update the numerator and denominator of the mixed energy estimator for a given iteration, one would have to compute the double sum $\sum_i w_i^{(t)} \sum_j H_{ij} t_j$. In order to speed up this calculation, we compute and store one of the two sums prior to the run:

$$N_i = \sum_{j \in \mathcal{T}} H_{ij} t_j.$$

We call these N_i the energy numerators (for consistency, we can also think in terms of the energy denominators, but these are no different from the trial wavefunction coefficients $D_i = t_i$). The number of energy numerators we have to store is $|C(\mathcal{T})|$, i.e., the number of determinants connected to the determinants in the trial wavefunction space $\mathcal{T}$.

Using the energy numerators and denominators, the mixed energy becomes

$$E_0^{\text{mixed}} = \frac{\sum_{t=1}^{N_{MC}} \sum_i w_i^{(t)} N_i}{\sum_{t=1}^{N_{MC}} \sum_i w_i^{(t)} D_i}.$$

Updating the mixed energy only requires finding which of the currently-occupied determnants overlap with stored energy numerators. This can be done by performing a binary search on the stored energy numerators for each currently occupied determinant, so updating the local energy each step takes only $\mathcal{O}\left(N_{walk} \ln N_{C(\mathcal{T})}\right)$ operations.

3.4.5 Alternative ideas for ψ_{T}

Thus, our deterministic space size is limited by the requirement that we be able to sparsely store the Hamiltonian connecting the deterministic states, while our trial wavefunction space size is limited by the requirement that we be able to store all determinants connected to them. In practice, this typically means that the deterministic space can be much larger than the trial wavefunction space, which leads us to consider alternative ideas for ψ_{T}, other than storing explicitly the coefficients of linear combinations of determinants ($|\psi_{\mathrm{T}}\rangle = \sum_{i \in \mathcal{T}} t_i |\phi_i\rangle$).

One such idea is the use of a matrix product state (MPS), which compactly encodes coefficients of all determinants in the Hilbert space. One can store two matrix product states, one representing $\left|\psi_{\mathrm{T}}^{\mathrm{MPS}}\right\rangle$ and one representing $H\left|\psi_{\mathrm{T}}^{\mathrm{MPS}}\right\rangle$. This means that the storage required to store an MPS trial wavefunction and its connections does not increase with the number of connections per determinant. Computing the overlap of an MPS with a given determinant can in theory be performed more quickly than computing the overlap of a given determinant with all connections to determinants enumerated in ψ_{T} represented as a linear combination. I therefore implemented an interface in our Fortran code that talks to C++ code from Garnet Chan's group that calculates MPS's. It has not been throroughly tested yet, since we ran into some bugs when we tried to go to larger Hilbert

spaces.

3.4.6 Modification: Graduated Initiator

Another change that we proposed with SQMC is the use of a "graduated initiator." Since we are "more sure" about the correctness of the signs of determinants that are in the deterministic space or close to it, we can apply an initiator threshold that changes with some measure of our "sureness" of sign-coherence. We choose to use the initiator cutoff

$$\text{init} = in^p,$$

where i and p are positive constants and n is the "distance from the deterministic space," defined as the least number of moves a walker on the determinant in question has taken since its last visit to the deterministic space.

It is not clear whether one should *a priori* expect the graduated initiator to be better or worse than the standard (constant) initiator. On the one hand, it seems logical, since the more steps a walker has taken from the deterministic space, the less certain we are of its correct sign. On the other hand, the graduated initiator makes the initiator threshold so high for states far from the deterministic space, that the walkers become confined to just a few steps away from the deterministic space, which should make the bias worse.

3.5 Results

The semistochastic method is now applied to compute the ground state energy of the carbon dimer and the simple-square 8×8 fermionic Hubbard model with

periodic boundaries. In both cases, we represent H in the basis of determinants formed from the restricted Hartree-Fock orbitals. For the Hubbard model these orbitals are the momentum eigenstates. For the carbon dimer these orbitals are obtained by solving the restricted Hartree-Fock equations in cc-pVTZ basis set [19]. The majority of the Hubbard calculations are performed for $U/t = 4$, where U is the on-site Coulomb repulsion and t is the nearest neighbor hopping parameter. This parametrization is considered to be in the *intermediate coupling* regime (the noninteracting bandwidth being $8t$), and has been used widely in the literature [3].

The trial wave function space and the deterministic space are generated with identical iterative schemes, but possibly different parameters. At each iteration, first define a reference space as all states obtained in the previous iteration. Second, generate a space which includes all determinants connected to the reference space by a single application of the Hamiltonian. Third, find the dominant eigenvector in this space. Fourth, truncate the space using a criterion based on the magnitude of the coefficient of each state in the eigenvector. This truncated space becomes the reference for the next iteration. The reference for the first iteration is the Hartree-Fock state.

For various sizes of the deterministic space, we demonstrate the improvements of SQMC over the purely stochastic method defined by a deterministic space which includes only the Hartree-Fock determinant. The purely stochastic method is almost the same as *i*-FCIQMC [11, 17], aside from some details such as the use of real walker weights versus the integer walker weights used in FCIQMC and the use of a graduated initiator in SQMC [4]. The most dramatic benefit of SQMC is in the efficiency, which is defined to be proportional to the inverse of the clock time required to obtain the ground state energy to a specified level of uncertainty.

To show the gain in efficiency of SQMC we computed the relative efficiency, *i.e.*, the efficiency normalized by that of the stochastic method ($|\mathcal{D}| = 1$), with $|\mathcal{T}| = 1$. Fig. 3.3 shows the relative efficiency of SQMC *vs.* the size of the deterministic space for the simple-square 8×8 Hubbard model with periodic boundaries, $U/t = 4$ and 10 electrons. The orders of magnitude increases in efficiency demonstrate the benefits not only of SQMC but also of improving the trial wave function. The gain of just using the largest deterministic space is a factor of 22, while the benefit of just using the largest trial wave function is a factor of 42. Both together yield a factor of about 900 as seen in the plot, but the two are not always multiplicative.

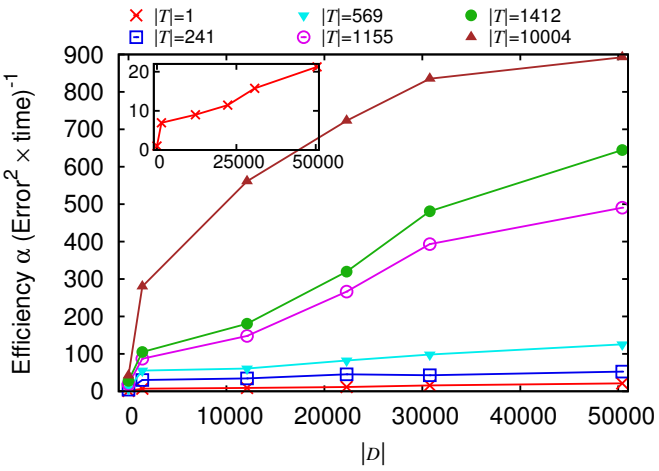

Figure 3.3: Relative efficiency of SQMC *vs.* dimension $|\mathcal{D}|$ of the deterministic space for the simple-square 8×8 Hubbard model with periodic boundaries, $U/t = 4$ and 10 electrons. Results are shown for trial wave functions of increasing size. The inset shows the $|\mathcal{T}| = 1$ curve on an expanded scale. For this system, $N \approx 10^{12}$.

Fig. 3.4 shows the efficiency gain of SQMC *vs.* filling fraction for the simple-square 8×8 Hubbard model with $U/t = 4$. The deterministic space, constructed by applying the Hamiltonian once to the Hartree-Fock determinant, has a rather modest increase in size from 1412 to 16540 determinants, whereas the size of the

Hilbert space grows enormously from about 10^{12} to 10^{35}. Nevertheless, the efficiency gains increase with filling fraction. Calculations beyond the scope of the present paper show that the initiator bias, at all fillings, decreases with increasing $\mathcal{D}$, but that it increases with filling fraction and U in both the stochastic and the semistochastic methods.

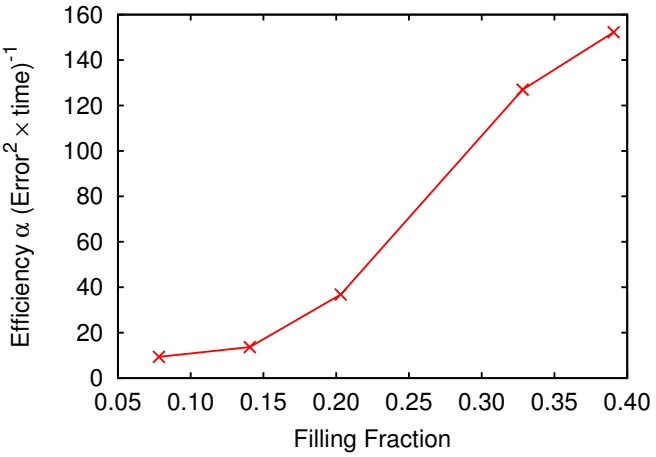

Figure 3.4: Relative efficiency of the SQMC *vs.* filling fraction for the simple-square 8×8 Hubbard model with $U/t = 4$. In all cases, the trial wave function is the Hartree-Fock determinant. The deterministic space is constructed by applying the Hamiltonian once to the Hartree-Fock determinant. This yields spaces of sizes 1412, 4088, 7424, 14160, 16540. N ranges from roughly 10^{12} to 10^{35}.

SQMC produces large efficiency gains for chemical systems as well. Fig. 3.5 shows the efficiency gain of SQMC *vs.* the size of the deterministic space for the carbon dimer with a cc-pVTZ basis set [19]. The bottom two curves are for $\mathcal{D}$ and $\mathcal{T}$ generated with one applications of our iterative scheme which generate single and double excitations only. The largest efficiency gain for these is about 40. The top two curves are for $\mathcal{D}$ and $\mathcal{T}$ generated with two applications of our iterative scheme and, hence, include several chemically relevant quadruple excitations which are important for correctly describing the ground state wave function. The largest

efficiency gain now jumps to over 1000.

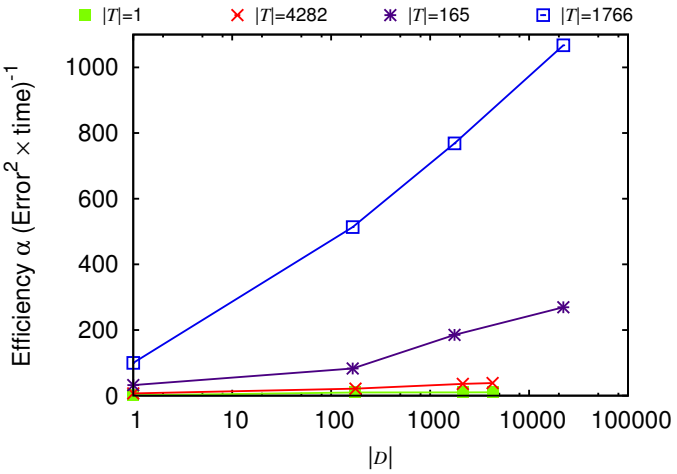

Figure 3.5: Relative efficiency of SQMC *vs.* dimension $|\mathcal{D}|$ of the deterministic space for the carbon dimer with a cc-pVTZ basis. Results are shown for trial wave functions of increasing size. The top two curves are for $\mathcal{D}$ and $\mathcal{T}$ generated with two applications of our iterative scheme. The 165 and 1766 determinant wave functions with some quadruple excitations have much higher efficiency than the 4282 determinant wave function without any. For this system, $N \approx 10^9$.

Not only is SQMC much more efficient than the stochastic method, but in some cases, also the initiator bias is significantly reduced. Fig. 3.6 shows the biased estimates of the energy as obtained by both the SQMC and stochastic method *vs.* the average number of occupied determinants for the 8×8 Hubbard model with $U/t = 1$ and 50 electrons. SQMC has essentially no bias. A larger average number of occupied determinants corresponds to using a larger walker population in the calculation. The time required for a calculation is proportional to the walker population.

The reduction in initiator bias is not always large. Fig. 3.7 shows both the SQMC and stochastic method energy *vs.* the average number of occupied determinants for the 8×8 Hubbard model with $U/t = 4$ and 10 electrons. SQMC

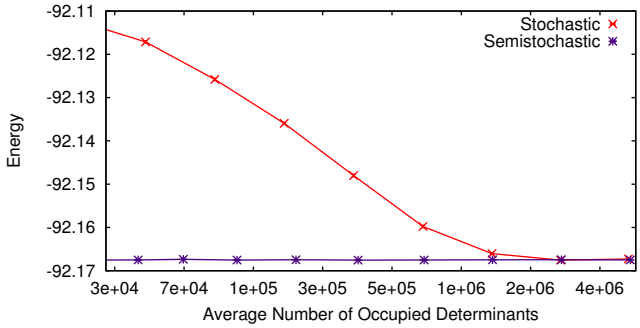

Figure 3.6: Energy of SQMC and the stochastic method *vs.* the average number of occupied determinants for the simple-square 8×8 Hubbard model with $U/t = 1$ and 50 electrons. The trial wave function for each of these calculations is the Hartree-Fock determinant. The deterministic space consists of the 16540 determinants connected to the Hartree-Fock determinant. For this system, $N \approx 10^{35}$.

has a reduced initiator bias for a small, but not for a large number of occupied determinants. However, for this system and all other systems studied SQMC has a smoother bias extrapolation curve than the stochastic method.

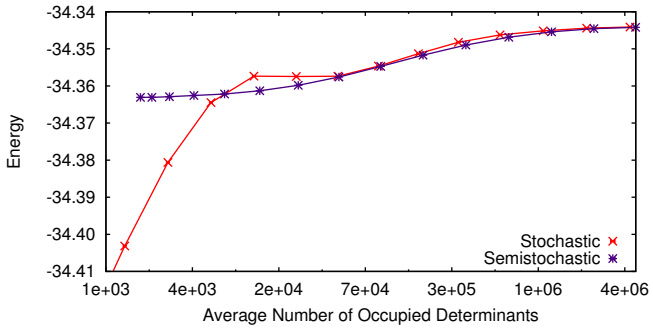

Figure 3.7: Energy of SQMC and the stochastic method *vs.* the average number of occupied determinants for the simple-square 8×8 Hubbard model with $U/t = 4$ and 10 electrons. The trial wave function for each of these calculations is the Hartree-Fock determinant. The deterministic space reference state for each SQMC calculation is the Hartree-Fock determinant, yielding a deterministic space of 1412 determinants. For this system, $N \approx 10^{12}$.

3.6 Conclusion

The semistochastic power method, a hybrid with deterministic and stochastic components, was introduced for finding the dominant eigenvalue and sampling the corresponding eigenvector of a matrix. We showed that this novel, deterministic component significantly reduces the noise of the purely stochastic method without compromising its ability to deal with matrices well beyond the size that can be handled by purely deterministic methods. In particular, matrices ranging in order from 10^9 to 10^{35} were successfully tackled. Besides being more efficient than a purely stochastic approach, the semistochastic method has in some cases the additional benefit of a much reduced initiator bias. Also, the bias tends to be smoother and more amenable to removal by extrapolation. We only presented applications to systems with a sign problem, but the efficiency benefits of a semistochastic implementation of the power method extend to systems without a sign problem.

3.7 Acknowledgments

We thank Garnet Chan and Ali Alavi for valuable discussions. This work was supported in part by DOE-CMCSN DE-SC0006650 (AAH), NSF CHE-1112097 (HJC) and NSF DMR-0908653 (FRP).

CHAPTER 4

EFFICIENT HEAT-BATH SAMPLING IN FOCK SPACE

In this chapter I introduce an algorithm for sampling many-body quantum states in Fock space. The algorithm efficiently samples states with probability approximately proportional to an arbitrary function of the second-quantized Hamiltonian matrix element connecting the sampled state to the current state. I apply the new sampling algorithm to the recently-developed Semistochastic Full Configuration Interaction Quantum Monte Carlo method (S-FCIQMC), a semistochastic implementation of the power method for projecting out the ground state energy in a basis of Slater determinants. Our new sampling method requires modest additional computational time and memory compared to uniform sampling but results in newly-spawned weights that are approximately of the same magnitude, thereby greatly improving the efficiency of projection. A comparison in efficiency between our sampling algorithm and uniform sampling is performed on the all-electron nitrogen dimer at equilibrium in Dunning's cc-pVXZ basis sets with $X \in \{\mathrm{D}, \mathrm{T}, \mathrm{Q}, 5\}$, demonstrating a large gain in efficiency that increases with basis set size. In addition, a comparison in efficiency is performed on three all-electron first-row dimers, B_2, N_2, and F_2, in a cc-pVQZ basis, demonstrating that the gain in efficiency compared to uniform sampling also increases dramatically with the number of electrons. This chapter was adapted from my recent paper published in the Journal for Chemical Theory and Computation [23].

4.1 Introduction

Methods for finding approximate solutions to the quantum many-body problem often work in Fock space, spanned by a discrete basis composed of products of N single-particle orbitals chosen from a set of M orbitals, where N is the number of particles. These products are symmetrized or antisymmetrized for bosonic or fermionic particles, respectively. However, the number of states in Fock space scales exponentially in M, and even the size of the sector of Fock space with constant particle number N scales combinatorially in M and N, so deterministic calculations are limited to small M and N.

When the number of states is too large for deterministic methods (approximately 10^{10}), Monte Carlo methods provide a valuable alternative. Many such methods have been developed in recent years. For example:

1. Full Configuration Interaction Quantum Monte Carlo (FCIQMC) [12, 18, 36, 10] and its semistochastic extension, denoted here as S-FCIQMC [31], have been used to calculate almost exact Full CI energies in far larger state spaces than deterministic methods can handle.
2. Density Matrix Quantum Monte Carlo (DMQMC) [5, 28] has been used to sample the density matrix using an S-FCIQMC-like dynamic.
3. Model Space Quantum Monte Carlo (MSQMC) [39], in which Slater determinants are stochastically sampled to generate an effective Hamiltonian using Löwdin partitioning for calculating excited states.
4. Monte Carlo Configuration Interaction (MCCI) [22], in which a variational wavefunction is calculated from a selected CI expansion generated by a Monte Carlo sampling procedure.

In all of these methods, Slater determinants are sampled *uniformly* from the set of determinants connected to a reference by the Hamiltonian. While uniform sampling has the advantage of being easy to implement, it is far from optimal, since some of the determinants connected to the reference are much more important than others. The definition of 'importance' here depends on the method used. In S-FCIQMC and DMQMC, starting from an initial state, the importance of a final state is the magnitude of the Hamiltonian matrix element connecting it to the initial state: if moves are proposed with probabilities proportional to this importance function, the weight of the final state is independent of which of the possible final states is chosen. In MCCI, the importance of a state could be the magnitude of the 2nd-order perturbation theory estimate of energy lowering achieved by including this state in the expansion. Since the importance of the states can range over many orders of magnitude, large gains in efficiency can be gained by sampling states in proportion to their importance. However, the naïve approach of computing the relative probabilities of all the connected states and then normalizing by dividing by their sum is prohibitively expensive because the number of connected states can be large. For example, for a quantum chemistry Hamiltonian, it scales as $\mathcal{O}\left(N^2M^2\right)$.

Motivated by the above, this chapter addresses the following problem: Given an initial configuration in Fock space, how can one *efficiently* sample new configurations with probability approximately proportional to a function of the Hamiltonian matrix element connecting the new configuration to the initial one? By 'efficient', we mean that the time complexity of one sampling event scales as $\mathcal{O}\left(N\right)$, the same time complexity as that of computing one Hamiltonian matrix element corresponding to a single excitation in a molecular system. We also limit our storage requirements to $\mathcal{O}\left(M^k\right)$, where k is the maximum number of creation and anni-

hilation operators in any off-diagonal term in the Hamiltonian − in other words, the same scaling as that required to store the integrals needed for computing the Hamiltonian. We take advantage of the fact that k is typically small; e.g., it is 4 for the quantum chemistry Hamiltonian and for the Hubbard model in momentum space, and it is 2 for the Hubbard model in real space. We demonstrate the efficiency gains achieved when the method is used in S-FCIQMC applied to the quantum chemistry Hamiltonian.

This chapter is organized as follows. Section 4.2 has an overview of the S-FCIQMC method and its application to the quantum chemistry Hamiltonian. In Section 4, our new sampling method is derived within the context of S-FCIQMC. In Section 5.4, the new method is compared to uniform sampling for S-FCIQMC calculations on the nitrogen dimer at equilibrium in several different basis sets, demonstrating that our sampling algorithm is more efficient than uniform sampling by a factor that increases rapidly with basis size. Similar calculations are also performed on the boron dimer and the fluoride dimer, demonstrating that at fixed number of orbitals, the gain in efficiency increases rapidly with electron number.

4.2 Background

4.2.1 Quantum chemistry Hamiltonian

The second-quantized nonrelativistic electronic Hamiltonian, in the Born-Oppenheimer approximation, is

$$\hat{H} = \sum_{pr} f_{rp} a_r^\dagger a_p + \frac{1}{2} \sum_{pqrs} g_{rspq} a_r^\dagger a_s^\dagger a_q a_p + h_{\text{nuc}}, \tag{4.1}$$

where $a_p^\dagger(a_p)$ is the usual electron creation (annihilation) operator with the indices $\{p,q,r,s\}$ incorporating both spatial and spin degrees of freedom. The tensors entering the expression of the Hamiltonian are (1) the 1-electron integrals,

$$f_{rp} = \int \phi_r^*(x)\left(-\frac{1}{2}\nabla^2 - \sum_I \frac{Z_I}{|\mathbf{r}-\mathbf{r}_I|}\right)\phi_p(x)\,dx, \tag{4.2}$$

with $\phi_p(x)$ denoting spin-orbitals, x the combined spatial ($\mathbf{r}$) and spin coordinates of the electrons, and Z_I and $\mathbf{r}_I$ the atomic number and spatial coordinates of nucleus I; (2) the 2-electron integrals,

$$g_{rspq} = \int \phi_r^*(x_1)\,\phi_s^*(x_2)\,\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}\phi_p(x_1)\,\phi_q(x_2)\,dx_1dx_2, \tag{4.3}$$

with an index-ordering convention according to the physicist notation [2]; and (3) the nuclear-nuclear repulsion,

$$h_{\text{nuc}} = \sum_{I<J}\frac{Z_I Z_J}{|\mathbf{r}_I-\mathbf{r}_J|}. \tag{4.4}$$

Orthogonal orbitals are used, and in the absence of a magnetic field can be chosen to be real in which case $f_{rp}=f_{pr}$ and $g_{rspq}=g_{psrq}=g_{rqps}=g_{pqrs}=g_{srqp}=g_{qrsp}=g_{spqr}=g_{qpsr}$. Hence the memory required for the 1-electron integrals is $N^2/2$ and that for the 2-electron integrals is $N^4/8$.

Since the Hamiltonian in Eq. 4.1 contains only 1-electron and 2-electron terms, it has nonzero matrix elements between two determinants only if they differ by no more than two spin-orbitals. Classifying the matrix elements by the number of spin-orbitals that the initial (i) and final states (f) differ in,

$$H_{\text{diag}} = \sum_{p\in\text{occ.}} f_{pp} + \frac{1}{2}\sum_{p,q\in\text{occ.}}(g_{pqpq}-g_{qppq}), \tag{4.5}$$

$$H(r\leftarrow p) = \Gamma^i_{rp}\left(f_{rp}+\sum_{q\in\text{occ.}}(g_{rqpq}-g_{qrpq})\right), \tag{4.6}$$

$$H(rs\leftarrow pq) = \Gamma^i_{rp}\Gamma^f_{sq}\Big(g_{rspq}-g_{srpq}\Big), \tag{4.7}$$

where $\Gamma^i_{rp} = (-1)^n$, n being the number of occupied spin-orbitals between p and r in state i.

Note that the magnitudes of the matrix elements for double excitations depend only on the four spin-orbitals whose occupations are changing, though the sign depends on the other occupied spin-orbitals as well. In contrast, the magnitudes of the diagonal and single-excitation matrix elements depend on all the occupied spin-orbitals.

4.2.2 S-FCIQMC overview

We now briefly review the Semistochastic Full Configuration Interaction Quantum Monte Carlo (S-FCIQMC) method [31], which is a modification of the FCIQMC method [12, 18, 36, 10]. There are three main differences between the S-FCIQMC method and the original FCIQMC method: namely, walkers have real rather than integer weights, the computation of the mixed estimator of the energy is done using a multideterminantal trial wavefunction rather than the Hartree-Fock determinant, and the part of the projection that involves the most important determinants is done deterministically rather than stochastically.

In common with all projector QMC methods, the S-FCIQMC method employs a *projector* operator, which is a function of the Hamiltonian such that the ground state of the Hamiltonian is the dominant state (state with largest absolute eigenvalue) of the projector. Repeated application of the projector to an arbitrary state that is not orthogonal to the ground state results in projecting onto the ground state. The S-FCIQMC method employs the linear projector,

$$\hat{G} \;=\; \hat{\mathbf{1}} + \tau \left(E_{\mathrm{T}} - \hat{H} \right), \tag{4.8}$$

where E_{T} is an estimate of the ground state energy. The time step τ must be smaller than $2/(E_{\max} - E_{\min})$, where $E_{\max}$ and $E_{\min}$ are the extremal eigenvalues of $\hat{H}$, in order for the ground state of $\hat{H}$ to be the dominant state of $\hat{G}$. In order to avoid negative diagonal matrix elements in $\hat{G}$, τ must be smaller than $1/[(H_{ii})_{\max} - E_{\min}]$. In practice, τ is chosen to roughly minimize the statistical error for given computer time, and this optimal value is yet smaller.

Since the total number of states is much larger than the number of states that can be stored on a computer, it is necessary to do at least part of the projection stochastically. Before the run, a subset of the determinants, dubbed the *deterministic space*, is selected. Projection between pairs of determinants that are both in the deterministic space is performed deterministically, using sparse matrix-vector multiplication. However, projection between a pair of determinants at least one of which is outside the deterministic space is performed stochastically, as follows.

At any Monte Carlo (MC) step t, the current state is represented as a sparse linear combination of Slater determinants that are currently "occupied," i.e.,

$$\left|\psi_0^t\right\rangle = \sum_{i \in \mathrm{occ.}} w_i^t \left|i\right\rangle, \tag{4.9}$$

where w_i^t is the weight on determinant $|i\rangle$ at step t. The state evolves from MC step t to step $t+1$ according to

$$\begin{aligned} w_j^{t+1} &= \sum_i \hat{G}_{ji} w_i^t \\ &= \sum_i \left[\delta_{ji} + \tau\left(E_{\mathrm{T}}\delta_{ji} - \hat{H}_{ji}\right)\right] w_i^t \\ &= \left(1 + \tau\left(E_{\mathrm{T}} - \hat{H}_{jj}\right)\right) w_j^t - \tau \sum_{i \neq j} \hat{H}_{ji} w_i^t. \end{aligned} \tag{4.10}$$

Note that the first of these terms is diagonal, and the second is off-diagonal. In addition to the off-diagonal elements connecting pairs of deterministic determinants,

all diagonal elements can be applied deterministically, since they do not increase the density of the sparse representation. The off-diagonal elements that do not connect pairs of deterministic states are sampled stochastically as follows.

On each iteration, the weight w_i^t on each Slater determinant i is divided up among an integer number n_i^t of *walkers*, where $n_i^t = \max(1, \lfloor |w_i^t| \rceil)$, and each walker spawns a new walker on determinant j with probability P_{ji} that receives a weight of

$$w_j^{(t+1)} = \frac{-\tau H_{ji}}{P_{ji}} \left(\frac{w_i^t}{n_i^t} \right). \tag{4.11}$$

To overcome the fermion sign problem, only determinants with absolute weight greater than an *initiator* threshold [18] are allowed to create weight on unoccupied determinants. The resulting bias in the energy disappears in the limit of infinite walker number. For efficiency reasons, walkers with weight less than a minimum weight are combined into a smaller number of larger weight walkers in a statistically unbiased fashion [31]. In this work, the minimum weight is set to 0.5, and a graduated initiator is used: the initiator threshold for a given determinant is equal to the minimum number of moves a walker on that determinant has taken since its last visit to the deterministic space.

An estimate of the ground state wavefunction could in principle be obtained by summing the walker distributions over all Monte Carlo iterations over a long run, i.e.,

$$|\psi_0\rangle \approx \sum_{t=1}^{N_{\mathrm{MC}}} |\psi_0^t\rangle = \sum_{t=1}^{N_{\mathrm{MC}}} \sum_{i \in \mathrm{occ.}} w_i^t |i\rangle . \tag{4.12}$$

However, the entire wavefunction $|\psi_0\rangle$ contains too many terms to be stored all at once; instead, only the sparse walker distribution $\{w_i^t\}$ at a given iteration can be stored. Therefore, to have an unbiased estimate of the energy, E_0 should

depend only linearly on the walker distribution. This is accomplished using a mixed estimator,

$$E_0^{\text{mix}} = \frac{\left\langle \psi_0 \left| \hat{H} \right| \psi_T \right\rangle}{\langle \psi_0 | \psi_T \rangle} = \frac{\sum_{t=1}^{N_{\text{MC}}} \sum_{i \in \text{occ.}} w_i^t N_i}{\sum_{t=1}^{N_{\text{MC}}} \sum_{i \in \text{occ.}} w_i^t t_i}, \tag{4.13}$$

where $|\psi_T\rangle = \sum_i t_i |i\rangle$ is a trial wavefunction that is chosen before the run, and $N_i = \sum_j H_{ij} t_j$ can also be computed and stored before the run. In the limit that either $|\psi_0\rangle$ or $|\psi_T\rangle$ is the exact ground state, the mixed energy is a zero-bias, zero-variance estimator of E_0. As a proxy for choosing $|\psi_T\rangle$ to be close to the true ground state, we choose it to be a low-energy linear combination of determinants.

Before starting an S-FCIQMC run, two sets of important determinants must be selected: the deterministic space and the determinants that make up the trial wavefunction expansion. In both important subspaces, including more determinants improves the efficiency, but the two subspaces have different storage constraints. The deterministic space must be small enough that the full Hamiltonian within the deterministic space can be stored (using compressed row storage) [1]. The number of determinants in $|\psi_T\rangle$ is limited by the requirement that all of the mixed energy numerators $N_i = \sum_j H_{ij} t_j$ can be stored. This means that the deterministic space can typically be much larger than the trial wavefunction expansion.

To select determinants for either subspace, the following iterative procedure is used. Starting with the Hartree-Fock determinant, all connected determinants are obtained. If the number of connected determinants is large, perturbation theory can be used to estimate the coefficient each determinant would have if a Lanczos diagonalization were performed; only the determinants with largest expected

[1] In our parallel implementation, the deterministic determinants are distributed among the cores. Each core must then have the rows of the projector corresponding to those determinants. In a serial run, only the nonzero elements in the upper-triangular part of the deterministic projector must be stored.

coefficients are retained. A Lanczos diagonalization is performed in the resulting space. Then, the determinants connected to the subset of determinants with the highest absolute coefficients are obtained, and the procedure is repeated. Finally, the determinant list is truncated to the required size, and in the case of the trial wavefunction, a final Lanczos diagonalization is performed.

Other methods for generating the trial wavefunction and deterministic space have also been proposed, including an automatic method that chooses important determinants based on the walker distribution after a short run [6]. It is possible that these methods may be advantageous in the present context, but comparing them is not the focus of this chapter.

4.2.3 Heat-bath distribution

The computed energy does not depend on P_{ji} in Eq. 4.11, provided that this is in fact the probability of proposing the move. However, the statistical error does depend on P_{ji}. Large weight fluctuations increase the statistical error in any Monte Carlo calculation, and their detrimental effect is even more severe when there is a sign problem. The *heat-bath* distribution is defined such that the proposal probability is proportional to the magnitude of the corresponding matrix element, i.e.,

$$P_{ji} = \frac{|H_{ji}|}{\sum_k |H_{ki}|}. \tag{4.14}$$

If the new states are chosen from the heat-bath distribution, then their weights are independent of which state is chosen.

The difficulty with heat-bath sampling is that it is prohibitively expensive to compute the full column sum $\sum_k |H_{ki}|$ every time an off-diagonal move is pro-

posed, since the number of off-diagonal elements is $\mathcal{O}(N^2M^2)$. Hence, until now, most S-FCIQMC calculations have been performed using an approximately uniform sampling of off-diagonal moves, which is computationally simpler since it only involves the number of off-diagonal elements − which can be calculated efficiently − rather than computing and summing them each iteration.

In this chapter, we introduce an efficient new approach for sampling an approximate heat-bath distribution, in which we factor the above proposal probability into probabilities of selecting each electron and unoccupied spin-orbital separately. These probabilities of individual steps can be computed and stored before the run in order to reduce the sampling time to $\mathcal{O}(N)$, which is the same scaling as that for computing one single-excitation matrix element. The algorithm is derived in the next section, and it is summarized in Appendix 4.A.

However, first we examine other approaches to proposing off-diagonal moves that had been tried in the literature prior to my Efficient Heat-bath Sampling algorithm.

4.3 Proposal of off-diagonal moves

In FCIQMC, the matrix-vector multiplication in the power method is performed stochastically, rather than deterministically: Rather than propagating the weight on one state to all states in the corresponding column of the projector, we divide up the weight into an integer number of walkers, and propagate each walker's weight onto just one other state connected to the current state by a nonzero projector matrix element (as well as deterministically propagating weight onto its current

state, using the diagonal projector matrix element).[2]

This leaves open the question of what distribution to use to propose off-diagonal moves. When a move is spawned with some proposal probability, the spawned weight is adjusted by dividing by that same probability, in order to keep the expectation values constant. Thus, as long as the set of moves assigned a nonzero proposal probability is unchanged, we are free to adjust the relative probabilities of those moves and still have a correct algorithm.

In SQMC, moves between pairs of states in the deterministic subspace are performed deterministically. In our current implementation, if an off-diagonal stochastic move is proposed that connects two states in the determininistic space, that move is discarded, since it is being done deterministically instead. Therefore, the use of a deterministic space does not complicate the process of generating the stochastic moves.

Since the magnitudes of the new weights are proportional to the corresponding off-diagonal projector matrix element, the ideal distribution to sample from is one in which the spawning probability is proportional to that off-diagonal element magnitude. This is referred to as the heat-bath distribution,

$$P\left(i \leftarrow j\right) = \frac{|H_{ij}|}{\sum_k |H_{kj}|}.$$

In heat-bath sampling, all newly spawned weights have the same magnitude, so we don't waste time on small-weight walkers. The unfortunate drawback is that heat-bath sampling requires computing the column sum in the denominator of the above expression, which can be extremely expensive in large systems with large numbers of nonzero off-diagonal elements per column.

[2]These off-diagonal stochastic propagation events are referred to interchangeably in these notes as "moving" or "spawning," even though "spawning" is the correct term, since the weight on the old determinant is not changed when new weight is created by it on another determinant.

We therefore seek a computationally efficient scheme for approximating the heat-bath distribution. Following are different options for sampling off-diagonal moves that attempt to do this, in both the Hubbard model and quantum chemistry, our two main application areas of SQMC.

4.3.1 Hubbard model

In an N-site Hubbard model with hopping parameter t and Coulomb parameter U with lattice spacing a, the momentum-space Hamiltonian is

$$\begin{aligned} H_{Hubbard} &= -2t \sum_{\vec{k},\sigma} \left[\cos k_x a + \cos k_y a\right] c^\dagger_{\vec{k}\sigma} c_{\vec{k}\sigma} \\ &\quad + \frac{U}{N} \sum_{\vec{k},\vec{q},\vec{Q}} c^\dagger_{\vec{k}-\vec{Q},\uparrow} c^\dagger_{\vec{q}+\vec{Q},\downarrow} c_{\vec{k},\uparrow} c_{\vec{q},\downarrow}, \end{aligned}$$

where $\vec{k}$ and $\vec{q}$ are the single-particle momenta ($(k_x, k_y) = \left(\frac{m\pi}{L_x}, \frac{n\pi}{L_y}\right)$), and $\vec{Q}$ represents the momentum transferred in the scattering event. The hopping term, which is proportional to t, is diagonal, while the scattering term, proportional to U, is off-diagonal. Note that the off-diagonal term only connects pairs of determinants that differ by only a single up-electron move and a single down-electron move, and only in a momentum-conserving way. Also note that nonzero off-diagonal elements have the same magnitude, so that in the space of momentum-conserving moves, heat-bath sampling is the same as uniform sampling. The following is our algorithm for generating an off-diagonal move:

1. Choose an up electron uniformly.
2. Choose a down electron uniformly.
3. Choose a momentum state not occupied by an up electron, uniformly.

4. Find what momentum the down electron must have after the move in order to conserve momentum.

5. If that momentum site is not already occupied by a down electron, move the up and down electrons to their new sites. Else, reject the move.

The proposal probability is then

$$P\left(i \to j\right) = \frac{1}{n_{\uparrow}} \times \frac{1}{n_{\downarrow}} \times \frac{1}{N_{sites} - n_{\uparrow}}.$$

The chance that a proposed move is discarded is

$$P\left(discard\right) = \frac{n_{\downarrow} - 1}{N_{sites} - 1},$$

since, aside from the one site that the down electron in question is located on, there are $N_{sites} - 1$ others, and only $n_{\downarrow} - 1$ of them are already occupied. This discard probability tends to be small for the filling fractions we typically run on (e.g., 8×8, $5 \uparrow$, $5 \downarrow$ has $P\left(discard\right) = 4/63 \approx 6\%$; 8×8, $13 \uparrow$, $13 \downarrow$ has $P\left(discard\right) = 12/63 \approx 19\%$).

4.3.2 Quantum Chemistry

The quantum chemistry Hamiltonian is given by

$$\hat{H} = \sum_{i} h\left(i\right) + \sum_{i<j} v\left(i,j\right) + V_{NN},$$

where the one-electron operator is

$$h\left(i\right) = -\frac{1}{2}\nabla_{i}^{2} - \sum_{A} \frac{Z_A}{r_{iA}},$$

the two-electron operator is

$$v\left(i,j\right) = \frac{1}{r_{ij}},$$

and V_{NN} is the nuclear-nuclear repulsion.

In the discrete Hilbert space of Slater determinants $\{|I\rangle\}$ composed of orthonormal orbitals $\{\chi_p(x_i)\}$, we can evaluate the matrix elements of the Hamiltonian $\left\langle I \left| \hat{H} \right| J \right\rangle$ by integrating products of orbitals and operators over the electron coordinates.

We define the one-electron integral as

$$(p\,|h|\,q) = \int dx_1 \chi_p^*(x_1)\, h(x_1)\, \chi_q(x_1),$$

and the two-electron integral as (using chemists' notation, as we do in our code)

$$(pq|rs) = \int dx_1 dx_2 \chi_p^*(x_1)\, \chi_q(x_1)\, \frac{1}{r_{12}} \chi_r^*(x_2)\, \chi_s(x_2).$$

For any two states $|I\rangle$ and $|J\rangle$ that differ by more than two orbitals, $\langle I\,|H|\,J\rangle = 0$, since all terms in it would involve an integral over at least one pair of orthonormal orbitals. The matrix elements are given by

$$\left\langle I \left| \hat{H} \right| J \right\rangle = \begin{cases} \sum_{p\in \text{OCC.}} (p\,|h|\,p) + \sum_{p,q\in \text{OCC.}} ((pp|qq) - (pq|qp)), & \text{diagonal element,} \\ \langle p\,|h|\,q\rangle + \sum_{r\in \text{OCC.}} ((pq|rr) - (pr|rq)), & \text{single excitation } (p \to q), \\ (pr|qs) - (ps|qr), & \text{double excitation } (pq \to rs), \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

In contrast to the Hubbard model, off-diagonal moves can be single excitation, double excitation with same spin, or double excitation with opposite spin. This means our move proposal algorithm now has to incorporate these three different types of moves.

Also, nonzero off-diagonal matrix elements of the quantum chemical Hamiltonian vary widely in magnitude. Therefore, in chemistry runs, we are left with a

choice in using heat-bath sampling, uniform sampling, or something in between. We consider each of these three options below.

Heat-bath sampling

Since ideally we would like to spawn weight of roughly the same magnitude, sampling would be performed most efficiently using the heat-bath distribution, where off-diagonal moves are chosen with probability proportional to the magnitude of the corresponding matrix element, i.e.,

$$P\left(i \leftarrow j\right) = \frac{|H_{ij}|}{\sum_k |H_{kj}|}.$$

This is the ideal distribution to sample from, since we don't waste time proposing moves via small off-diagonal elements that will end up spawning a small weight. It also has the added bonus that all three types of moves are generated with correct relative probabilities.

Unfortunately, for any reasonable basis set size, it is prohibitively expensive to compute the column sum $\sum_k |H_{kj}|$ every time an off-diagonal move is proposed, since the number of off-diagonal elements is $\mathcal{O}\left(n_{el}^2 N_{orb}^2\right)$.

Uniform sampling

The idea of uniform sampling is to choose from the list of nonzero off-diagonal moves with equal probability. This has the advantage that generating moves is very fast, but the disadvantage that it is far from the ideal heat-bath distribution, so it can spawn weights of widely varying magnitudes.

(Note that our implementation of "uniform" sampling is not exactly uniform,

as will be seen in what follows. However, our goal was not to exactly sample the uniform distribution, but to quickly sample some distribution, approximately uniform or otherwise, so this is not an issue.)

To generate a move, we must decide whether to make a single excitation, double excitation with same spin, or double excitation with opposite spin. We do this by enumerating the number of possible moves of each type as follows:

$$
\begin{aligned}
n_{single} &= n_{\uparrow}\left(N_{orb} - n_{\uparrow}\right) + n_{\downarrow}\left(N_{orb} - n_{\downarrow}\right), \\
n_{double}^{\uparrow\uparrow} &= \frac{n_{\uparrow}\left(n_{\uparrow} - 1\right)}{2} \times \frac{\left(N_{orb} - n_{\uparrow}\right)\left(N_{orb} - n_{\uparrow} - 1\right)}{2}, \\
n_{double}^{\downarrow\downarrow} &= \frac{n_{\downarrow}\left(n_{\downarrow} - 1\right)}{2} \times \frac{\left(N_{orb} - n_{\downarrow}\right)\left(N_{orb} - n_{\downarrow} - 1\right)}{2}, \\
n_{double}^{\uparrow\downarrow} &= n_{\uparrow} n_{\downarrow}\left(N_{orb} - n_{\uparrow}\right)\left(N_{orb} - n_{\downarrow}\right).
\end{aligned}
$$

Note that these are actually upper bounds since symmetry is being ignored, but the ratios between them would not change much if the fraction of moves that are disallowed by symmetry is approximately constant for each of these types of moves. This is the first source of non-uniformity in our routine.

We pick a single excitation with probability

$$
P\left(single\right) = \frac{n_{single}}{n_{single} + n_{double}^{\uparrow\uparrow} + n_{double}^{\downarrow\downarrow} + n_{double}^{\uparrow\downarrow}}
$$

and otherwise pick a double excitation. Now that we have determined the number of electrons to excite, we use one of the two following algorithms to actually generate the excitations.

Single excitations

This algorithm is straight-forward.

1. Choose one electron uniformly. It can be of either spin.
2. If there is a non-empty set of orbitals of the correct symmetry that are not already occupied by an electron of that spin, choose one uniformly and move the electron there. Else, reject the move.

This yields the simple proposal probability

$$P = \frac{1}{n_\sigma} \times \frac{1}{N_{\text{symmetry-allowed unoccupied}_\sigma \text{ orbitals}}}.$$

As can be seen clearly from this expression, the proposal probability is not constant, since the number of symmetry-allowed unoccupied orbitals will not in general be the same for all chosen electrons. This is the second source of non-uniformity.

As a simple example of how this is non-uniform, consider a system with 5 orbitals and 2 up-electrons, where the first 3 orbitals have one symmetry and orbitals 4 and 5 have another. Consider the determinant with only orbitals 1 and 4 occupied. This determinant should make 3 off-diagonal moves with equal probability:

$$P(1 \to 2) = P(1 \to 3) = P(4 \to 5) = \frac{1}{3}.$$

However, using this algorithm, the probabilities would be

$$\begin{aligned} P(1 \to 2) = P(1 \to 3) &= \frac{1}{4}, \\ P(4 \to 5) &= \frac{1}{2}, \end{aligned}$$

which is clearly not uniform.

Double excitations

We can perform a double excitation in a manner similar to how we did for Hubbard. In chemistry, our moves must now follow point-group symmetry conservation, rather than momentum conservation, and in general there is more than one orbital of a given symmetry label. Therefore, once one electron's destination is chosen, there is usually more than one orbital that the other electron is allowed to move into.

1. Choose two electrons uniformly. They can be of either spin. If they are opposite spin, arbitrarily label the up electron as electron 1 and the down electron as electron 2.
2. Choose an orbital not occupied by an electron of electron 1's spin, uniformly.
3. Find what symmetry electron 2 must have after the move in order to obey symmetry laws.
4. If there is a non-empty set of orbitals of the correct symmetry that are not already occupied by an electron with the same spin as electron 2, choose one uniformly and move both electrons. Else, reject the move.

The proposal probability for opposite-spin moves is then

$$P^{\uparrow\downarrow} = \frac{2}{n_{el}\left(n_{el}-1\right)} \times \frac{1}{N_{orb}-n_{\uparrow}} \times \frac{1}{N_{\text{symmetry-allowed unoccupied}\downarrow\text{ orbitals}}},$$

where the factor of 2 comes from the fact that the two electrons can be chosen in either order. For same-spin moves it is

$$P^{\sigma\sigma} = \frac{2}{n_{el}\left(n_{el}-1\right)} \times \frac{1}{N_{orb}-n_{\sigma}} \times \left(P\left(\text{hole 1}|\text{hole 2}\right) + P\left(\text{hole 2}|\text{hole 1}\right)\right),$$

where $P\,(\text{hole } j|\text{hole } i)$ is the probability of choosing hole j given that hole i was already chosen uniformly. This probability is

$$P\,(\text{hole } j|\text{hole } i) = \frac{1}{N_{\text{symmetry-allowed unoccupied } \downarrow \text{ orbitals after hole } i \text{ was already chosen}}},$$

and in general, $P\,(\text{hole } j|\text{hole } i) \neq P\,(\text{hole } i|\text{hole } j)$. This inequality hints at why this probability is again non-uniform.

As an example of how this algorithm produces non-uniform moves, consider the same system as in the previous example, with 5 orbitals, the first 3 of which have one symmetry and the last two have another symmetry. Consider the determinant with orbitals 1 and 4 occupied by up electrons, and now add down electrons at orbitals 1 and 2. Assume that moving a pair of electrons of the first symmetry to the second symmetry is allowed. Now let's consider just the opposite-spin double excitations. There are three, which should have equal probability:

$$P^{\uparrow\uparrow}\left((1,4) \to (2,5)\right) = P^{\uparrow\uparrow}\left((1,4) \to (3,5)\right) = P^{\downarrow\downarrow}\left((1,2) \to (4,5)\right) = \frac{1}{3}.$$

However, the algorithm as written above yields instead

$$\begin{aligned} P^{\uparrow\uparrow}\left((1,4) \to (2,5)\right) = P^{\uparrow\uparrow}\left((1,4) \to (3,5)\right) &= \frac{1}{4}, \\ P^{\downarrow\downarrow}\left((1,2) \to (4,5)\right) &= \frac{1}{2}, \end{aligned}$$

which again is not uniform. This is a third source of non-uniformity.

Lack of actual uniformity

Note that our "uniform" spawning is not truly uniform for several (overlapping) reasons:

1. As mentioned before, we choose whether to perform a single excitation or a double excitation with probability based on the combinatorics *without considering symmetry*. For some determinants, the proportion of symmetry-allowed moves that are single excitations is very different from the proportion of all moves (symmetry-allowed or -disallowed) that are single excitations.
2. In all three types of off-diagonal moves (single, same-spin double, and opposite-spin double), the proposal probability is a function of the number of unoccupied orbitals of each symmetry group, and therefore is not constant.

The combination of these factors could easily lead the relative probabilities of "uniform" off-diagonal moves to differ from one another by an order of magnitude or more, especially for systems with a large basis set.

For example, with Be_2 in a $v4z$ basis with $D_\infty h$ symmetry, there is one symmetry group that contains 26 orbitals and another that contains only 2. By analogy with the single-excitation example I gave above, which had 3 orbitals of one symmetry group and only 2 of another, there are cases where the relative probabilities of single-excitation moves are $\frac{1}{2} : \frac{1}{50}$, which is already different by a factor of 25!

4.3.3 Cauchy-Schwarz sampling

The Alavi group has recently developed a new distribution that roughly approximates the ideal heat-bath distribution without being much more expensive than uniform. It is based on the Cauchy-Schwarz inequality $|\vec{u} \cdot \vec{v}| \leq |\vec{u}|\,|\vec{v}|$.

The two-body integrals

$$(ik|jl) = \int dx_1 dx_2 \chi_i^* (x_1) \chi_k (x_1) \frac{1}{r_{12}} \chi_j^* (x_2) \chi_l (x_2)$$

can be bounded in magnitude using Cauchy-Schwarz as

$$|(ik|jl)| \leq \sqrt{(ik|ik)(jl|jl)}.$$

The magnitudes of off-diagonal matrix elements can then be approximated as

$$\begin{aligned} \left|H^{off}((i,j) \to (k,l))\right| &= |(ik|jl) - (il|jk)| \\ &\leq \sqrt{(ik|jl)^2 - (il|jk)^2} \\ &\leq \sqrt{(ik|jl)^2 + (il|jk)^2} \\ &\leq \sqrt{(ik|ik)(jl|jl) + (il|il)(jk|jk)}, \end{aligned}$$

or the following looser bound, which is easier to use:

$$\begin{aligned} \left|H^{off}((i,j) \to (k,l))\right| &= |(ik|jl) - (il|jk)| \\ &\leq |(ik|jl)| + |(il|jk)| \\ &\leq \sqrt{(ik|ik)(jl|jl)} + \sqrt{(il|il)(jk|jk)} \end{aligned}$$

In what follows, we use "Cauchy-Schwarz sampling" to mean heat-bath sampling with the magnitudes of off-diagonal elements approximated by converting the above inequality into an equality:

$$\left|H_{CS}^{off}((i,j) \to (k,l))\right| = \sqrt{(ik|ik)(jl|jl)} + \sqrt{(il|il)(jk|jk)}$$

We propose the following algorithms for generating single and double excitations. We will deal with the probability of choosing whether to propose a single or double excitation later, but the following are proposed algorithms for generating off-diagonal moves of a given excitation level:

Double excitations

Before the run, store $\sqrt{(ij|ij)}$ for all pairs $i \neq j$. Also, for each orbital i and symmetry label s, compute and store

$$G_s(i) = \sum_{j \neq i, \mathrm{sym}(j)=s} \sqrt{(ij|ij)}.$$

Also, store

$$S_2(i) = \sum_s G_s(i).$$

During the run, given a determinant, propose off-diagonal moves corresponding to double excitations as follows:

1. For each occupied orbital i, compute

$$S_2'(i) = \sum_{j \in unocc.} \sqrt{(ij|ij)} = S_2(i) - \sum_{j \neq i,\, j \in occ.} \sqrt{(ij|ij)}.$$

2. Choose electron i with probability

$$P(i) = \frac{S_2'(i)}{\sum_{k \in occ.} S_2'(k)}.$$

3. Choose electron j with probability

$$P(j|i) = \frac{S_2'(j)}{\sum_{k \in occ.} S_2'(k) - S_2'(i)}.$$

4. At this point, we have selected two electrons, i and j, but we could have selected them in either order. Therefore, record the probability of selecting them as

$$P(i,j) = P(i)\,P(j|i) + P(j)\,P(i|j).$$

With probability $\frac{1}{2}$, reverse the labels of electrons i and j. This is necessary because up until now, we have assumed that i and j are interchangeable, but in the following, we spawn with electron i first.

5. Choose an unoccupied orbital k with the same spin as electron i with probability

$$P\left(i \rightarrow k\right) = \frac{\sqrt{\left(ik|ik\right)}}{S_2'\left(i\right)}.$$

6. Choose an unoccupied orbital of the same spin as electron j with the correct symmetry s, with probability

$$P\left(j \rightarrow l | i \rightarrow k\right) = \frac{\sqrt{\left(jl|jl\right)}}{G_s\left(j\right) - \sum_{m \neq j,\, \mathrm{sym}(m)=s} \sqrt{\left(jm|jm\right)}}.$$

   If no such orbitals exist, no move is proposed.

7. Since these two unoccupied orbitals could have been selected in either order,

$$\begin{aligned} P\left(\left(i,j\right) \rightarrow \left(k,l\right) | i,j\right) \;=\; & \frac{1}{2} \times \left[P\left(i \rightarrow k\right) P\left(j \rightarrow l | i \rightarrow k\right)\right. \\ & + P\left(i \rightarrow l\right) P\left(j \rightarrow k | i \rightarrow l\right) \\ & + P\left(j \rightarrow k\right) P\left(i \rightarrow l | j \rightarrow k\right) \\ & \left. + P\left(j \rightarrow l\right) P\left(i \rightarrow k | j \rightarrow l\right)\right] \end{aligned}$$

   where the last two terms are needed because we could have spawned with electron j first rather than with electron i. The second and third terms in the square brackets are 0 in the case of opposite-spin excitations. The final probability of spawning a double excitation is

$$P\left(\left(i,j\right) \rightarrow \left(k,l\right)\right) = P\left(i,j\right) P\left(\left(i,j\right) \rightarrow \left(k,l\right) | i,j\right).$$

Note that the Alavi group is currently using a simpler algorithm where the pair of electrons is chosen uniformly, i.e., $P\left(i,j\right) = \frac{2}{n_{el}(n_{el}-1)}$.

Single excitations

The single excitation matrix element of a move from orbital i to orbital j is given by

$$H\left(i \rightarrow j\right) = \left(i\left|h\right|j\right) + \sum_{k \in occ.} \left(\left(ij|kk\right) - \left(ik|kj\right)\right).$$

Since this involves a sum over the currently occupied orbitals, it is not something that we can precompute, like we can with double excitations. Therefore, I'm looking for an approximation. I'm hoping that we can drop the second term, since it is much more time-consuming to compute. It could be the case that on average the second term is close to zero, since it is the difference between two positive quantities. If that is the case, then it would just be a noise term on the probability distribution, but we should check to make sure it is the case.

Before the run, compute and store the absolute values of the matrix elements of

$$H'\left(i \rightarrow j\right) = \left(i\left|h\right|j\right)$$

as well as

$$S_1\left(i\right) = \sum_{j \neq i} \left|H'\left(i \rightarrow j\right)\right|.$$

During the run, for a given determinant, propose off-diagonal moves corresponding to single excitations as follows:

1. For each occupied orbital i, compute

$$S_1'\left(i\right) = \sum_{j \in unocc.} \left|H'\left(i \rightarrow j\right)\right| = S_1\left(i\right) - \sum_{j \neq i,\, j \in occ.} \left|H'\left(i \rightarrow j\right)\right|.$$

2. Choose electron i with probability

$$P\left(i\right) = \frac{S_1'\left(i\right)}{\sum_{k \in occ.} S_1'\left(k\right)}.$$

3. Choose an unoccupied orbital of the same spin and symmetry with probability

$$P\left(i \to j|i\right) = \frac{\left|H'\left(i \to j\right)\right|}{\sum_{k \in unocc.,\, sym} \left|H'\left(i \to k\right)\right|},$$

where the sum is over unoccupied orbitals of the same symmetry as orbital i, if any exist. Else, don't propose a move.

4. The probability of spawning is

$$P\left(i \to j\right) = P\left(i\right) P\left(i \to j|i\right).$$

Choosing between single and double excitations

Now that we have investigated methods for proposing moves of a given excitation level (single or double), consider how to determine the relative probabilities of choosing the two excitation levels. We should choose with probability proportional to the estimated sum of absolute-value off-diagonal elements for each excitation level, which I'll call M_1 and M_2 for single and double excitations, respectively.

Recall that for single excitations, an electron is chosen with probability

$$P\left(i\right) = \frac{S_1'\left(i\right)}{\sum_{k \in occ.} S_1'\left(k\right)}.$$

This came from the fact that we estimated the sum of magnitudes of off-diagonal elements as the denominator of that expression, i.e.,

$$M_1 \approx \sum_{k \in occ.} S_1'\left(k\right).$$

For double excitations, a pair of electrons (i, j) is chosen with probability

$$\frac{S_2'\left(i\right)}{\sum_{k \in occ.} S_2'\left(k\right)} \times \frac{S_2'\left(j\right)}{\sum_{k \in occ.} S_2'\left(k\right) - S_2'\left(i\right)} + \left(i \leftrightarrow j\right).$$

Following similar logic, since the denominator is the product of two sums over products of $\sqrt{(ij|ij)}$ terms, and since we are assuming the approximation

$$\left|H^{off}\left((i,j)\rightarrow(k,l)\right)\right| \approx \sqrt{(ik|ik)\,(jl|jl)} + \sqrt{(il|il)\,(jk|jk)},$$

the denominator should be approximately the sum of the magnitudes of off-diagonal elements. We can approximate it as follows:

$$\begin{aligned} M_2 &\approx \sum_{k,k'\in occ.;k<k'} S_2'(k)\, S_2'(k') \\ &= \frac{1}{2}\left[\left(\sum_{k\in occ.} S_2'(k)\right)^2 - \sum_{k\in occ.} \left(S_2'(k)\right)^2\right]. \end{aligned}$$

Given these estimates for the sums of magnitudes of off-diagonal matrix elements, we choose single excitations with probability

$$P(single) = \frac{M_1}{M_1 + M_2},$$

and choose double excitations otherwise.

4.4 Approximate heat-bath algorithm

The nonrelativistic quantum chemistry Hamiltonian contains off-diagonal terms corresponding to single and double excitations only. Double excitations are more numerous, but much simpler to deal with than single excitations. While the number of double excitations from a given determinant is $\mathcal{O}(N^2M^2)$, the number of distinct values of double excitation matrix elements in the full Hamiltonian is only $\mathcal{O}(M^4)$, and each double excitation matrix element takes only $\mathcal{O}(1)$ time to compute. On the other hand, while the number of single excitations possible from a given determinant is only $\mathcal{O}(NM)$, the number of distinct values of single excitation

matrix elements in the full Hamiltonian is combinatorially large, and each matrix element takes $\mathcal{O}(N)$ time to compute, since the matrix elements for exciting from a given spin-orbital include a sum over all other occupied spin-orbitals (see Eq. 4.6). So, efficient computation of single-excitation matrix elements is difficult because there are too many to store and they are expensive to compute on the fly.

Since double excitations account for most of the possible excitations from a given determinant, and since they are so much easier to deal with efficiently, we start by constructing an approximate heat-bath algorithm for double excitations, and then we incorporate single excitations into the algorithm.

4.4.1 Sampling double excitations

If single excitations are ignored, heat-bath sampling requires us to sample moves from spin-orbitals $\{p, q\}$ to spin-orbitals $\{r, s\}$ with probability

$$P(rs \leftarrow pq) = \frac{|H(rs \leftarrow pq)|}{\sum_{p'q' \in \text{occ.}} \sum_{r's' \in \text{unocc.}} |H(r's' \leftarrow p'q')|}, \tag{4.16}$$

where the sum over $\{p', q'\}$ is over all occupied spin-orbitals, and the sum over $\{r', s'\}$ is over all unoccupied spin-orbitals.

We can factor this probability into a four-step process as follows:

$$\begin{aligned} P(rs \leftarrow pq) &= \left(\frac{\sum_{q' \in \text{occ.}} \sum_{r's' \in \text{unocc.}} |H(r's' \leftarrow pq')|}{\sum_{p'q' \in \text{occ.}} \sum_{r's' \in \text{unocc.}} |H(r's' \leftarrow p'q')|} \right) \\ &\quad \times \left(\frac{\sum_{r's' \in \text{unocc.}} |H(r's' \leftarrow pq)|}{\sum_{q' \in \text{occ.}} \sum_{r's' \in \text{unocc.}} |H(r's' \leftarrow pq')|} \right) \\ &\quad \times \left(\frac{\sum_{s' \in \text{unocc.}} |H(rs' \leftarrow pq)|}{\sum_{r's' \in \text{unocc.}} |H(r's' \leftarrow pq)|} \right) \left(\frac{|H(rs \leftarrow pq)|}{\sum_{s' \in \text{unocc.}} |H(rs' \leftarrow pq)|} \right) \end{aligned} \tag{4.17}$$

$$\equiv P(p)\ P(q|p)\ P(r|p,q)\ P(s|p,q,r)\,. \tag{4.18}$$

The conditional probabilities of selecting unoccupied spin-orbitals for double excitation, $P(r|p,q)$ and $P(s|p,q,r)$, will be different for each determinant, since they involve sums over all the spin-orbitals that are currently unoccupied. However, we can approximate them by summing over *all* spin-orbitals, except for p and q. Similarly, in the probability for selecting the first occupied spin-orbital, $P(p)$, the sum on q' over occupied orbitals can be replaced by a sum over all orbitals except orbital p. Defining

$$D_{pq} = \sum_{r's' \notin \{p,q\}} |H(r's' \leftarrow pq)| \quad \text{and} \quad S_p = \sum_{q' \neq p} D_{pq'}, \tag{4.19}$$

the approximate heat-bath probabilities are:

$$\begin{aligned}
\tilde{P}(rs \leftarrow pq) &= \left(\frac{\sum_{q' \neq p} \sum_{r's' \notin \{p,q'\}} |H(r's' \leftarrow pq')|}{\sum_{p' \in \text{occ.}} \sum_{q' \neq p'} \sum_{r's' \notin \{p',q'\}} |H(r's' \leftarrow p'q')|} \right) \\
&\quad \times \left(\frac{\sum_{r's' \notin \{p,q\}} |H(r's' \leftarrow pq)|}{\sum_{q' \in \text{occ.}} \sum_{r's' \notin \{p,q'\}} |H(r's' \leftarrow pq')|} \right) \\
&\quad \times \left(\frac{\sum_{s' \notin \{p,q,r\}} |H(rs' \leftarrow pq)|}{\sum_{r's' \notin \{p,q\}} |H(r's' \leftarrow pq)|} \right) \left(\frac{|H(rs \leftarrow pq)|}{\sum_{s' \notin \{p,q,r\}} |H(rs' \leftarrow pq)|} \right) \\
&\equiv \frac{S_p}{\sum_{p' \in \text{occ.}} S_{p'}} \frac{D_{pq}}{\sum_{q' \in \text{occ.}} D_{pq'}} \tilde{P}(r|p,q) \; \tilde{P}(s|p,q,r)
\end{aligned} \tag{4.20}$$

$$\equiv \tilde{P}(p) \; \tilde{P}(q|p) \; \tilde{P}(r|p,q) \; \tilde{P}(s|p,q,r). \tag{4.21}$$

Since S_p, D_{pq}, $\tilde{P}(r|p,q)$, and $\tilde{P}(s|p,q,r)$ all involve sums over *all* spin-orbitals, they can be computed and stored once at the beginning of the run. However, $\tilde{P}(p)$ and $\tilde{P}(q|p)$ involve sums over only the currently-occupied spin-orbitals, and thus they must be computed on the fly (in $\mathcal{O}(N)$ time). The unoccupied spin-orbitals can be sampled from $\tilde{P}(r|p,q)$, and $\tilde{P}(s|p,q,r)$ in $\mathcal{O}(1)$ time using the alias method [42], as described in Appendix 4.D.

Note that approximating the heat-bath probabilities affects only the efficiency of the calculation, not its exactness. Once the initiator bias has been extrapolated

away, the results are exact (i.e., they have a statistical error but no bias), as long as the same probabilities are used for proposing moves and computing the weights.

4.4.2 Sampling single excitations

We now modify the double-excitation proposal algorithm above to include the possibility of proposing a single excitation. After selecting electrons in spin-orbitals $\{p, q\}$ and empty spin-orbital r, we allow two possibilities:

1. Select a second unoccupied spin-orbital s, generating the double excitation $(rs \leftarrow pq)$ (as described above), OR
2. Discard the already-selected occupied spin-orbital q and generate the single excitation $(r \leftarrow p)$.

The heat-bath distribution requires us to choose between a single and a double excitation with probability proportional to their corresponding matrix elements:

$$\tilde{P}\left(\text{single}|p,q,r\right) \propto |H(r \leftarrow p)| \quad \text{and} \quad \tilde{P}\left(\text{double}|p,q,r\right) \propto H^{\text{tot}}_{rpq}, \tag{4.22}$$

where

$$H^{\text{tot}}_{rpq} \equiv \sum_{s' \notin \{p,q,r\}} |H(rs' \leftarrow pq)| \tag{4.23}$$

is the sum of double-excitation off-diagonal elements in which electrons in spin-orbitals p and q excite to two *other* spin-orbitals, one of which is r. Note that H^{tot}_{rpq} can be precomputed and stored at the beginning of the run.

One method of choosing between single and double excitations is to choose a single excitation with probability $\frac{|H(r \leftarrow p)|}{H^{\text{tot}}_{rpq} + |H(r \leftarrow p)|}$ and to choose a double excitation

otherwise. However, the problem with this approach is that if $|H(r \leftarrow p)| \gg H^{\text{tot}}_{rpq}$ the double excitation probability becomes small, so when a double excitation is chosen, the newly-spawned walker gets a large weight. A solution to this problem is to make both a double-excitation and a single-excitation move if $|H(r \leftarrow p)| > H^{\text{tot}}_{rpq}$. Using this approach, the probabilities of selecting a single or double excitation are

$$\tilde{P}\left(\text{single}|p,q,r\right) = \begin{cases} \frac{|H(r \leftarrow p)|}{H^{\text{tot}}_{rpq}+|H(r \leftarrow p)|}, & \text{if } |H(r \leftarrow p)| < H^{\text{tot}}_{rpq}; \\ 1, & \text{otherwise,} \end{cases} \tag{4.24}$$

and

$$\tilde{P}\left(\text{double}|p,q,r\right) = \begin{cases} \frac{H^{\text{tot}}_{rpq}}{H^{\text{tot}}_{rpq}+|H(r \leftarrow p)|}, & \text{if } |H(r \leftarrow p)| < H^{\text{tot}}_{rpq}; \\ 1, & \text{otherwise,} \end{cases} \tag{4.25}$$

respectively. Now, there can still be some large-weight single-excitation moves when $|H(r \leftarrow p)| \gg H^{\text{tot}}_{rpq}$, but since single-excitation moves are a very small fraction of all moves, this is not a serious problem.

4.4.3 Symmetry considerations

One of the advantages of our algorithm is that spatial symmetry-violating moves are automatically disallowed, since their corresponding matrix elements are zero. Time-reversal symmetry can also be straightforwardly included, just as with uniform sampling [8], as described in Appendix 4.E.

4.4.4 Check for correctness of heat-bath sampling

While all valid double excitations can be proposed using this algorithm, one may wonder whether the same is true of single excitations. The only situation in which

a valid single excitation $(r \leftarrow p)$ cannot be proposed by our heat-bath algorithm is if all other occupied spin-orbitals $q \neq p$ are the only ones of their irreducible representation, since then there would be no symmetry-allowed double excitation $(rs \leftarrow pq)$, and $\tilde{P}\,(r|p,q)$ would be zero. Hence, a sufficient − but not necessary − condition for heat-bath sampling to be correct for a given system consisting of $n_\uparrow$ spin-up electrons and $n_\downarrow$ spin-down electrons is:

$$\max\left(n_\uparrow, n_\downarrow\right) \;\; > \;\; N'_{\text{irrep.}}, \tag{4.26}$$

where $N'_{\text{irrep.}}$ is the number of irreducible representations consisting of only one spatial orbital.

Care must be taken that the full symmetry is used in this step. It is common practice to use D_{2h} symmetry, rather than the full $D_{\infty h}$ symmetry, for a homonuclear diatomic molecule. For example, suppose there is only one orbital pair that transforms as the Δ_g irreducible representation of $D_{\infty h}$. The member of the pair that transforms as $x^2 - y^2$ belongs to the A_g irreducible representation of D_{2h} whereas the member that that transforms as xy belongs to the B_{1g} irreducible representation of D_{2h}. However, there are other functions, 1, $x^2 + y^2$ and z^2 that also transform as the A_g representation. So, using the irreducible representations of $D_{\infty h}$ we would conclude that this orbital pair contributes two to $N'_{\text{irrep.}}$ whereas using the irreducible representations of D_{2h} symmetry we would incorrectly conclude that this orbital pair contributes one to $N'_{\text{irrep.}}$.

In practice, when we do not know the irreducible representations of the orbitals in the full symmetry group, we calculate $N'_{\text{irrep.}}$ as follows. Consider the system consisting of the same set of M spatial orbitals but only one electron. Compute the Full CI Hamiltonian matrix H' (dimension M) for this system. The number of columns of H' that have no nonzero off-diagonal elements is equal to $N'_{\text{irrep.}}$.

This check is done at the beginning of a run to ensure that heat-bath sampling is unbiased for the system being examined.

4.4.5 Alternative algorithm to avoid check for correctness

Since single excitations account for such a small fraction of all off-diagonal Hamiltonian elements ($\mathcal{O}(NM)$ singles compared to $\mathcal{O}(N^2M^2)$ doubles), it doesn't make much difference whether single excitations are proposed using heat-bath sampling. One possibility is to first decide whether to spawn a single or double excitation with some probability p_{sing}, which can be dynamically updated during the run to make the average spawned weight approximately the same for singles and doubles. Then, double excitations should be spawned using the heat-bath probabilities above (with $\tilde{P}(\text{double}|p,q,r) = 1$), while single excitations could be sampled uniformly.

If this alternative scheme of spawning single excitations is used, the check for correctness in Section 4.4.4 is no longer needed.

4.5 Results: B_2, N_2 and F_2 molecules

The relative efficiency of the approximate heat-bath method to the uniform method was tested on the B_2, N_2 and F_2 molecules. Dunning's cc-pVQZ basis was used for all three molecules, and in addition, for N_2, cc-pVXZ bases with $X \in \{\text{D}, \text{T}, \text{Q}, 5\}$ were used to study the basis set dependence. For each basis set, the determinants in both the deterministic space and the trial wavefunction were selected from the set of determinants that are at most quadruple excitations from the Hartree-Fock determinant. The size of the deterministic space was 4×10^4 determinants, and the

trial wavefunction contained 10^3 determinants (except for cc-pV5Z, which had a trial wavefunction with only 500 determinants). To accelerate convergence, time-reversal symmetry (see Appendix 4.E) and natural orbitals from a second-order Moller-Plesset perturbation theory (MP2) calculation were used. The natural orbitals and electron integrals were generated using the MOLPRO software package [43].

Fig. 4.1 shows the frequency of spawned absolute weights (in units of τ) for the uniform and the approximate heat-bath methods applied to the N_2 molecule at equilibrium geometry in cc-pVDZ and cc-pV5Z basis sets. In both basis sets, the vast majority of the absolute weights spawned by the heat-bath method lie in a narrow range, whereas those spawned by the uniform method range over orders of magnitude.

The reduction in the spread of the spawned weights makes the initiator approximation [18] more meaningful and allows one to use a larger time step τ. In the uniform sampling method, a state can become an initiator simply because it happened to receive a large weight from a single determinant. In the absence of an initiator, the S-FCIQMC method yields unbiased energies, provided that $\tau < 2/(E_{\max} - E_{\min})$. When a nonzero initiator threshold is used, there is an initiator bias which depends on the time step. Since the initiator limits the spawning more at small τ than at large τ, one may expect that the initiator bias would stay about the same or go down with increasing τ (although the statistical error would eventually go up because of the sign problem), both for uniform sampling and for heat-bath sampling. In our tests, we found that switching from uniform to heat-bath sampling had a negligible effect on the initiator bias for a given number of walkers.

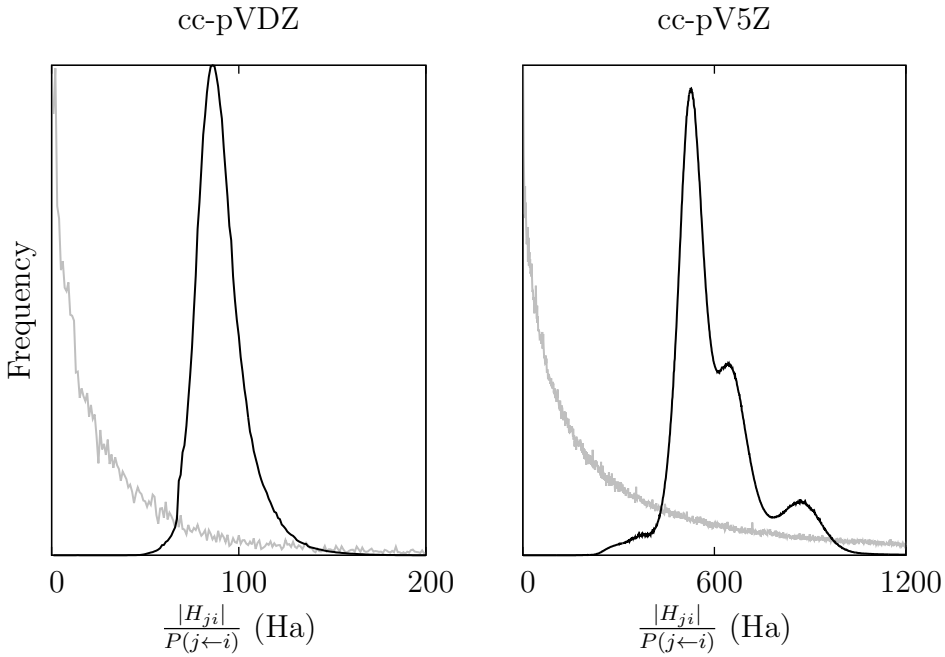

Figure 4.1: Probability distributions of off-diagonal absolute weights (in multiples of τ), spawned by the uniform (gray) and heat-bath (black) algorithms for equilibrium N_2, in cc-pVDZ and cc-pV5Z basis sets with all electrons correlated and 10^6 walkers. In the new heat-bath sampling algorithm, the spawned off-diagonal weights are close in magnitude, whereas in uniform sampling, the values of the spawned weights span many orders of magnitude. At optimal τ, the locations of the peaks in both heat-bath distributions correspond to off-diagonal weights of about 0.13, while the average spawned weights were about 0.15. Each histogram was accumulated during the course of a whole run and included approximately 10^{10} spawning events.

The increase in τ speeds up the time needed to achieve equilibration, as well as increasing the efficiency of the method. Since the statistical uncertainty, σ_E, in the estimate of the ground state energy decreases as the inverse square root of the Monte Carlo run time T, the efficiency is defined to be

$$\text{Efficiency} \propto \frac{1}{\sigma_E^2 T}. \tag{4.27}$$

We note that unlike the diffusion Monte Carlo method where the equilibration

time is a negligible part of the total time, in S-FCIQMC the equilibration can be a substantial part of the total.

For the cc-pV5Z basis, Fig. 4.2 shows how the relative efficiencies change with τ. We denote the value of τ that maximizes the efficiency as the *optimal* value, τ_{opt}. The relative efficiencies are scaled such that the peak for uniform sampling is at unity. Not only is the peak efficiency 32 times higher for heat-bath sampling, but the peak is broader, which makes it easier to choose a value of τ that is close to optimal. In this work, we found the optimal values of τ for both heat-bath and uniform sampling by doing several runs with different time steps to find which yields the greatest efficiency. However, we note that for heat-bath sampling, for all the systems studied here, a nearly-optimal value of τ can be obtained by choosing it such that the average magnitude of spawned weight is about 0.15. This scheme is preferable, as it avoids having to do many runs just to choose an optimal τ.

Table 4.1 shows these quantities for a sequence of three molecules. The efficiency gain increases with the number of electrons from 3.8 for B_2, to 54 for F_2. For uniform sampling, when N or M is large, it is hard to get an accurate estimate of the statistical error because (1) the optimal τ is very small, making the autocorrelation time in units of the number of Monte Carlo steps very long, and (2) in both sampling methods the population must be sufficiently large to enable cancellations to occur. Thus, in Tables 4.1 and 4.2, the equilibration speedup and increase in optimal time step are given to only single-digit accuracy.

In a sign problem-free method, one would expect that the efficiency would increase linearly with τ in the limit of small τ. However, in S-FCIQMC, within each sampling method, the severity of the sign problem also increases with τ. This is why the highest efficiency for a given sampling method is not achieved near

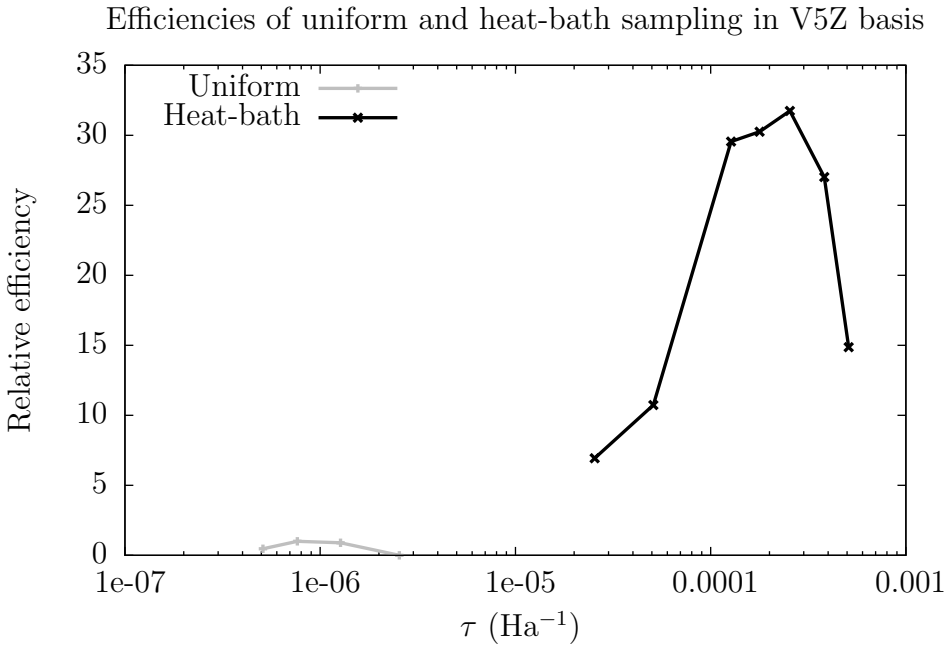

Figure 4.2: Comparison of relative efficiency vs. time step size for uniform and heat-bath methods on the all-electron nitrogen dimer at equilibrium in the cc-pV5Z basis. Efficiency plotted relative to the greatest efficiency seen for uniform. Heat-bath sampling improves efficiency by a factor of 32, while the optimal time step is about a factor of 300 larger.

$1/\left[(H_{ii})_{\max} - E_{\min}\right]$. It is also the reason that the increase in efficiency obtained upon switching from uniform sampling to heat-bath sampling is less than the increase in τ_{opt}.

molecule	efficiency gain	equilibration speedup	τ_{opt} increase
B_2	3.8	5	20
N_2	31	30	200
F_2	54	50	200

Table 4.1: Efficiency gains for all-electron equilibrium calculations of the ground state energy in a cc-pVQZ basis for different molecules. The efficiency gain increases with the number of electrons.

Table 4.2 demonstrates that the efficiency gain increases with the size of the

basis. For the N_2 molecule, the gain increases from 3.4 for the cc-pVDZ basis to 32 for the cc-pV5Z basis. We note that in our implementation, heat-bath sampling is about 1.8 times slower than uniform sampling for cc-pV5Z, so with a more efficient implementation, the gain in efficiency could be significantly larger.

basis	efficiency gain	equilibration speedup	τ_{opt} increase
cc-pVDZ	3.4	2	5
cc-pVTZ	16	30	200
cc-pVQZ	31	30	200
cc-pV5Z	32	60	300

Table 4.2: Efficiency gains and the factor by which the optimal time step τ_{opt} increases upon using heat-bath sampling instead of uniform sampling, for N_2 at the equilibrium geometry including core excitations. For a given sampling algorithm, τ_{opt} is the time step that maximizes the efficiency. The gain in efficiency of heat-bath sampling over uniform sampling increases with increasing basis set size.

In summary, the gain in efficiency from using heat-bath sampling rather than uniform sampling increases both with the number of electrons and the number of spin-orbitals.

4.6 Discussion

The efficient heat-bath sampling algorithm introduced in this chapter enables one to sample many-body quantum states connected to an initial state with probability approximately proportional to an arbitrary function of the Hamiltonian matrix element connecting the two states. Efficient heat-bath sampling has the potential to be useful in any of the Fock-space Monte Carlo methods, such as S-FCIQMC [12, 18, 36, 10, 31], DMQMC [5, 28], MSQMC [39], MCCI [22], and even Fock-space

variational Monte Carlo (VMC) [16].

We chose the probabilities to be proportional to the corresponding matrix element, which greatly improves the efficiency of S-FCIQMC. However, other choices are possible; e.g., in MCCI, it would be useful to choose the probabilities to be proportional to the lowering in energy estimated from second-order perturbation theory.

Finally, we note that the gain in efficiency from using the heat-bath method will be even greater for excited states if they are calculated using a projector that involves sampling the square of the Hamiltonian [7], since products of weights fluctuate even more than the weights themselves.

Acknowledgments: We thank Frank Petruzielo, Alessandro Roggero, Matt Otten, and George Booth for contributing to the development of our S-FCIQMC code, and Miguel Morales and George Booth for valuable discussions. This work was supported by grants NSF CHE-1112097, DOE DE-SC0006650, and NSF ACI-1534965. After moving to UIUC, H.J.C. was supported by the SciDAC program under DOE Award Number DE-FG02-12ER46875.

4.A Recap of algorithm when applied to S-FCIQMC

Before the S-FCIQMC run, compute and store the quantities S_p, D_{pq}, $\tilde{P}\left(r|p,q\right)$, $\tilde{P}\left(s|p,q,r\right)$, and H^{tot}_{rpq}, as described in Appendix 4.C.

When spawning off-diagonal moves from determinant $|i\rangle$, first compute and store the probability distribution for selecting the first occupied spin-orbital p

from this determinant,

$$\tilde{P}(p) = \frac{S_p}{\sum_{p'} S_{p'}}, \tag{4.28}$$

and compute the A and Q arrays needed for the alias method (see Appendix 4.D). This step takes $\mathcal{O}(N)$ time.

Divide up the weight w_i on $|i\rangle$ into $n_i = \max(1, \lfloor |w_i| \rceil)$ walkers. Then, for each walker on determinant $|i\rangle$, spawn one or two new determinants as follows:

1. Using the alias method, choose the first occupied spin-orbital p from the stored distribution $\tilde{P}(p)$. This step takes $\mathcal{O}(1)$ time.
2. Compute the probability distribution for choosing the second occupied spin-orbital q given that p was already selected,

$$\tilde{P}(q|p) = \frac{D_{pq}}{\sum_{q'} D_{pq'}}, \tag{4.29}$$

   and sample the second occupied spin-orbital q from $\tilde{P}(q|p)$. This step takes $\mathcal{O}(N)$ time.
3. Choose a spin-orbital $r \notin \{p, q\}$ from the stored distribution $\tilde{P}(r|p, q)$, such that p and r are of the same spin, using the alias method. If r is occupied in $|i\rangle$, no moves are generated by this walker; go to next walker. Otherwise, r is the first unoccupied spin-orbital for this excitation. This step takes $\mathcal{O}(1)$ time.
4. Decide whether to spawn a single excitation, a double excitation or one of each, as follows. Compute the single-excitation matrix element $H(r \leftarrow p)$. If $|H(r \leftarrow p)|$ is greater than the stored quantity H^{tot}_{rpq}, then spawn both the single excitation $(r \leftarrow p)$ and a double excitation; otherwise spawn only one

of the two types of excitations as follows. With probability

$$\frac{|H(r \leftarrow p)|}{H^{\text{tot}}_{rpq} + |H(r \leftarrow p)|}, \tag{4.30}$$

choose the single excitation $(r \leftarrow p)$; otherwise choose a double excitation. This step takes $\mathcal{O}(N)$ time, since that is the complexity of computing a single-excitation matrix element.

5. If a double excitation is to be proposed, choose a fourth spin-orbital $s \notin \{p, q, r\}$ from the stored distribution $\tilde{P}(s|p, q, r)$, using the alias method. If s is occupied in $|i\rangle$, no double excitation is generated by this walker. Otherwise, s is the second unoccupied spin-orbital for this excitation, and the move $(rs \leftarrow pq)$ is generated. This step takes $\mathcal{O}(1)$ time.

6. After generating the excitation(s), compute the weight spawned on the new determinant(s) as follows. If a single excitation $(r \leftarrow p)$ was proposed to a new determinant $|j\rangle$, the weight spawned on $|j\rangle$ is

$$w_j^{(t+1)} = \frac{-\tau H(r \leftarrow p)}{\tilde{P}(r \leftarrow p)} \left(\frac{w_i^t}{n_i}\right), \tag{4.31}$$

where $\tilde{P}(r \leftarrow p)$ is given in Appendix 4.B. If a double excitation $(rs \leftarrow pq)$ was proposed to a new determinant $|k\rangle$, the weight spawned on $|k\rangle$ is

$$w_k^{(t+1)} = \frac{-\tau H(rs \leftarrow pq)}{\tilde{P}(rs \leftarrow pq)} \left(\frac{w_i^t}{n_i}\right), \tag{4.32}$$

where $\tilde{P}(rs \leftarrow pq)$ is given in Appendix 4.B.

4.B Proposal probabilities

As discussed in the text, when a new walker is spawned, the weight assigned to that walker is proportional to the corresponding matrix element of the projector

divided by the proposal probability for that transition. The computation of these proposal probabilities is described next. The proposal probabilities both for single and double excitations can be computed in $\mathcal{O}(N)$ time.

4.B.1 Single excitations

The probability for a single excitation $(r \leftarrow p)$ is

$$\tilde{P}(r \leftarrow p) = \sum_{q' \in \text{occ.}} \tilde{P}(p)\, \tilde{P}(q'|p)\, \tilde{P}(r|p,q')\, \tilde{P}(\text{single}|p,q',r), \tag{4.33}$$

where

$$\tilde{P}(\text{single}|p,q,r) = \begin{cases} \frac{|H(r\leftarrow p)|}{H^{\text{tot}}_{rpq}+|H(r\leftarrow p)|}, & \text{if } |H(r \leftarrow p)| < H^{\text{tot}}_{rpq}; \\ 1, & \text{otherwise} \end{cases} \tag{4.34}$$

is the probability of choosing the single excitation $(r \leftarrow p)$ given that spin-orbitals $\{p,q,r\}$ have already been selected.

An alternative probability is $\tilde{P}'(r \leftarrow p) = (N-1)\,\tilde{P}(p)\,\tilde{P}(q|p)\,\tilde{P}(r|p,q)$, where q is the spin-orbital that was selected and discarded during the heat-bath sampling routine. As described in Appendix 4.F, this alternative method has an efficiency tradeoff associated with it: it increases efficiency by avoiding having to compute all terms in the sum, but it also decreases efficiency by increasing the fluctuations in spawned weights. This alternative probability was not employed in this chapter.

4.B.2 Double excitations

The probability for a double excitation $(rs \leftarrow pq)$ is

$$\begin{aligned}\tilde{P}\left(rs \leftarrow pq\right) &= \tilde{P}\left(p\right)\tilde{P}\left(q|p\right)\Big[\tilde{P}\left(r|p,q\right)\tilde{P}\left(\text{double}|p,q,r\right)\tilde{P}\left(s|p,q,r\right) \\ &\qquad + \tilde{P}\left(s|p,q\right)\tilde{P}\left(\text{double}|p,q,s\right)\tilde{P}\left(r|p,q,s\right)\Big] \\ &\quad + \tilde{P}\left(q\right)\tilde{P}\left(p|q\right)\Big[\tilde{P}\left(r|q,p\right)\tilde{P}\left(\text{double}|q,p,r\right)\tilde{P}\left(s|q,p,r\right) \\ &\qquad + \tilde{P}\left(s|q,p\right)\tilde{P}\left(\text{double}|q,p,s\right)\tilde{P}\left(r|q,p,s\right)\Big],\end{aligned} \tag{4.35}$$

where

$$\tilde{P}\left(\text{double}|p,q,r\right) = \begin{cases} \frac{H^{\text{tot}}_{rpq}}{H^{\text{tot}}_{rpq}+|H(r\leftarrow p)|}, & \text{if } |H(r\leftarrow p)| < H^{\text{tot}}_{rpq}; \\ 1, & \text{otherwise} \end{cases} \tag{4.36}$$

is the probability of choosing a double excitation given that $\{p,q,r\}$ have already been selected.

The second and third terms in Eq. 4.35 are zero for opposite-spin excitations.

As with single excitations, an alternative probability is

$$\tilde{P}'\left(rs \leftarrow pq\right) = c\tilde{P}\left(p\right)\tilde{P}\left(q|p\right)\tilde{P}\left(r|p,q\right)\tilde{P}\left(\text{double}|p,q,r\right)\tilde{P}\left(s|p,q,r\right), \tag{4.37}$$

where c is either 2 or 4 for opposite-spin or same-spin double excitations, respectively. Again, this alternative probability was not employed in this chapter.

4.C Quantities that must be precomputed at start of run

At the start of the run, compute and store the following tensors. It is assumed that up- and down-spin orbitals are the same, but the number of up- and down-spin

electrons need not be the same.

1. Electron pair selection probabilities tensor:

$$D_{pq} = \sum_{r's'} |H(r's' \leftarrow pq)| \tag{4.38}$$

for all $p \neq q$. This has size $2M(2M-1)/2$, since electrons can be of either spin. The storage could be reduced to $M(M-1)/2$ for same-spin electrons and M^2 for opposite-spin electrons.

2. Single electron selection probabilities tensor:

$$S_p = \sum_{q' \neq p} D_{pq'} \tag{4.39}$$

for all p. The electrons are chosen one at a time with probabilities $\tilde{P}(p)$ and $\tilde{P}(q|p)$ computed using S_p and D_{pq}, respectively. S has size M.

3. First hole selection probabilities tensor:

$$\tilde{P}(r|p,q) = \frac{\sum_{s'} |H(rs' \leftarrow pq)|}{\sum_{r's'} |H(r's' \leftarrow pq)|} \tag{4.40}$$

for all $p \neq q \neq r$. This is the probability of choosing the first empty spin-orbital r, given that electrons in spin-orbitals p and q have already been selected for excitations. This can be stored as two separate tensors, $\tilde{P}_{\text{same}}(r|p,q)$ and $\tilde{P}_{\text{opposite}}(r|p,q)$, for same-spin and opposite-spin double excitations, respectively. The same-spin excitation tensor only has to be of size $\frac{1}{2}M^3$, while the opposite-spin excitation tensor can be of size M^3. In that case, $\tilde{P}_{\text{opposite}}(r|p,q)$ represents the probability of choosing r given p,q with r and p having same spin.

4. Double excitations probabilities tensor:

$$\tilde{P}(s|p,q,r) = \frac{|H(rs \leftarrow pq)|}{\sum_{s'} |H(rs' \leftarrow pq)|} \tag{4.41}$$

for all $p \neq q \neq r \neq s$. When sampling a double excitation, this is the probability of choosing the second empty spin-orbital s given that electrons in spin-orbitals p and q and empty spin-orbital r have already been chosen. Also, store the denominators $H^{\text{tot}}_{rpq} = \sum_{s'} |H(rs' \leftarrow pq)|$, the summed magnitudes of double excitation matrix elements. This will be needed for comparing with single excitation matrix elements. The tensor can be stored as two separate tensors, $\tilde{P}_{\text{same}}(s|p,q,r)$ and $\tilde{P}_{\text{opposite}}(s|p,q,r)$, of sizes $\frac{1}{2}M^4$ and M^4, respectively.

For each of the two probability tensors $\tilde{P}(r|p,q)$ and $\tilde{P}(s|p,q,r)$, we store not only the probabilities, but the corresponding A and Q tensors for sampling them in $\mathcal{O}(1)$ time using the alias method (see Appendix 4.D). Thus, the total storage space for $\tilde{P}(s|p,q,r)$ and its corresponding A and Q tensors is $\frac{3}{2}M^4$ integers and $3M^4$ single-precision real numbers. In comparison, the integrals files that are already being stored are $\mathcal{O}\left(\frac{1}{8}M^4\right)$ double-precision real numbers, so the storage requirement is 18 times larger than the storage requirement for the integrals alone.

4.C.1 Relaxing the storage requirements

While the above storage requirements have the same $\mathcal{O}(M^4)$ scaling as the two-body integrals that are already stored, the fact that $\tilde{P}(s|p,q,r)$ requires 18 times the storage of the integrals can become prohibitive for more than about 250 orbitals. One possibility is to estimate the 2-body integrals, for example using an approximation inspired by the Cauchy-Schwarz inequality [37]. Instead, here we discuss a method that avoids storing $\tilde{P}(s|p,q,r)$ for all $\mathcal{O}(M^4)$ combinations $\{p,q,r,s\}$ without making the Cauchy-Schwarz approximation.

Note that $\tilde{P}(s|p,q,r)$ is only sampled when $\{p,q,r\}$ have already been selected. Some $\{p,q,r\}$ combinations occur a lot more often than others, so $\tilde{P}(s|p,q,r)$ needs to be stored only for the most frequently-occurring triplets $\{p,q,r\}$. For the other, less frequently-occurring triplets $\{p,q,r\}$, one could either choose spin-orbital s uniformly, or compute the correct heat-bath conditional probability $\tilde{P}(s|p,q,r)$ on the fly in $\mathcal{O}(M)$ time.

As can be seen in Fig. 4.3, the fraction of $\{p,q,r\}$ triplets that account for the vast majority of calls to sample $\tilde{P}(s|p,q,r)$ decreases with basis set size. In order to account for 99% of the number of calls to sample $\tilde{P}(s|p,q,r)$, for cc-pVDZ, one must store $\tilde{P}(s|p,q,r)$ for about 50% of the $\{p,q,r\}$ triplets, while for cc-pV5Z, it is only required to store $\tilde{P}(s|p,q,r)$ for about 20% of the triplets.

In addition to this reduction in the storage requirements, one need only store the nonzero elements of the 4-index tensors. The largest basis used in this chapter is cc-pV5Z (182 orbitals), but with these modifications, calculations with a cc-pV6Z basis (280 orbitals) could easily be done.

4.D Alias method for sampling from discrete distributions

Discrete distributions consisting of M probabilities can be sampled straightforwardly by constructing an array of cumulative probabilities, drawing a random number, and doing a binary search to find the interval in which the random number falls. This requires $\mathcal{O}(M)$ time to set up the cumulative probabilities and $\mathcal{O}(\log M)$ time to sample.

However, when the same distribution is being sampled repeatedly, the alias

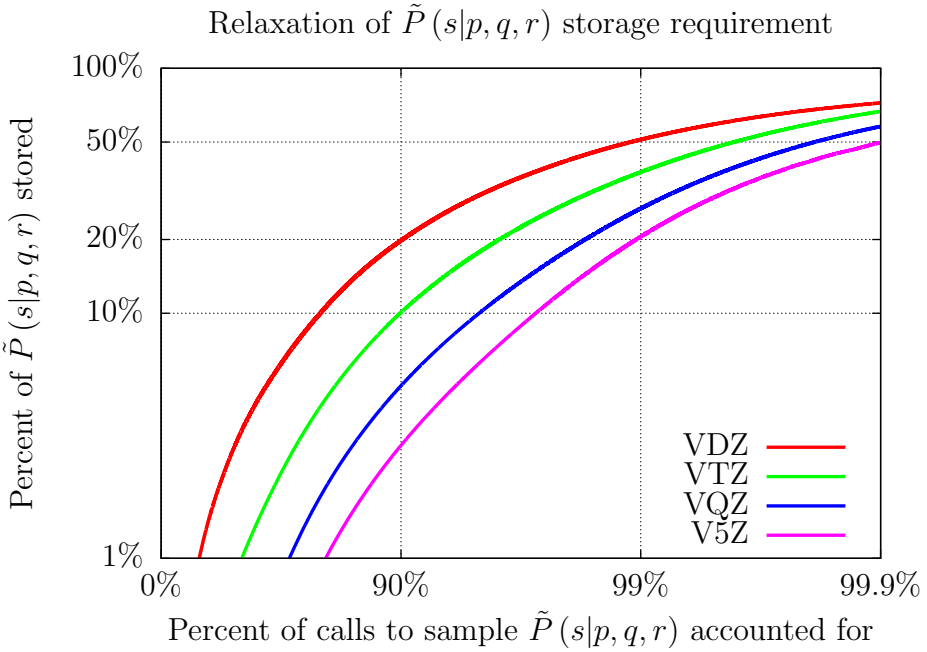

Figure 4.3: This plot shows what percentage of $\tilde{P}(s|p,q,r)$ must be stored, in order for the stored part of $\tilde{P}(s|p,q,r)$ to account for a given percentage of the total calls to sample it, for an all-electron nitrogen dimer in cc-pVXZ basis sets for $X \in \{D,T,Q,5\}$. It was obtained by counting the number of times that each triplet $\{p,q,r\}$ was selected (followed by sampling $\tilde{P}(s|p,q,r)$) during a run with 10^5 walkers.

method [42] can be used to sample it in even more quickly, in $\mathcal{O}(1)$ time. The method employs a real array, Q, and an integer array A, each of length M. Sampling with the alias method is done as follows. First, orbital i is selected uniformly from the set of all M possible orbitals. Then, with probability Q_i, orbital i is sampled; else, orbital A_i is sampled. The cost of doing this is just the cost of drawing two uniform random numbers.

The set of probabilities $\{Q_i\}$ and aliases $\{A_i\}$ can be generated in $\mathcal{O}(M)$ time at the beginning of the run [27], as shown in the pseudocode in Fig. 4.4.

4.E Time-reversal symmetry

The size of the Hilbert space can be reduced by almost a factor of two when the number of up and down electrons are equal by taking advantage of time-reversal symmetry [8]. This increases the effectiveness of cancellations, and allows one to use larger deterministic spaces and trial wavefunctions in S-FCIQMC. It also enables one to calculate an excited state as easily as the ground state provided that it is the lowest state of a different symmetry under time reversal than the ground state; e.g., if the ground state is a singlet, it enables calculating the lowest triplet state. All calculations in this chapter make use of time-reversal symmetry and here we discuss details of how it is implemented in S-FCIQMC.

Let $\hat{T}$ denote the time-reversal operator, so $|i'\rangle \equiv \hat{T}|i\rangle$ is the state obtained by flipping all the spins of the electrons in state $|i\rangle$. Since $\hat{T}^2|i\rangle = |i\rangle$, the eigenvalues of the time-reversal operator, z, must be ± 1. Even S states have $z = 1$, and odd S states have $z = -1$, where S is the total spin of the system. So, the wavefunctions can be expanded in a symmetrized basis,

$$|\tilde{i}\rangle = \frac{1}{\sqrt{2}N_i}\left(|i\rangle + z|i'\rangle\right), \tag{4.42}$$

with normalization factor

$$N_i = \begin{cases} 1, & \text{if } \langle i|i'\rangle = 0, \\ \sqrt{2}, & \text{otherwise.} \end{cases} \tag{4.43}$$

Note that a state which is its own time-reversed partner ($|i\rangle = |i'\rangle$) can only have non-zero coefficients in a wavefunction that has $z = 1$.

In each symmetrized linear combination $|\tilde{i}\rangle$, one of the states is chosen to be the *representative*, $|i\rangle_{\text{rep}}$. The representative has a positive coefficient in the linear

combination, and non-representative state has the same coefficient multiplied by z. In Eq. (4.42) $|i\rangle_{\text{rep}} = |i\rangle$. The representative is chosen by converting the computer representation of states (a binary string) into a number and then defining

$$|i\rangle_{\text{rep}} \equiv \min\left(|i\rangle, |i'\rangle\right). \tag{4.44}$$

Thus the representative is the state that comes first according to an arbitrary (but consistent) convention for ordering.

Recall that when time-reversal symmetry is not used, a walker that makes an off-diagonal move from state $|i\rangle$ to state $|j\rangle$ is assigned weight, $w_j = \frac{-\tau H_{ji}}{P_{ji}}\left(\frac{w_i}{n_i}\right)$, where w_i and n_i are the weight and number of walkers on $|i\rangle$, respectively. This viewpoint is adopted even with the inclusion of time-reversal symmetry, albeit with modifications to the Hamiltonian matrix element and the proposal probability, to give

$$w_{\tilde{j}} = -\frac{\tau \tilde{H}_{\tilde{j}\tilde{i}}}{P_{\tilde{j}\tilde{i}}}\left(\frac{w_{\tilde{i}}}{n_{\tilde{i}}}\right), \tag{4.45}$$

where $\tilde{H}_{\tilde{j}\tilde{i}}$ is the Hamiltonian matrix element between the two symmetrized pairs of states, $\tilde{i}$ and $\tilde{j}$, and $P_{\tilde{j}\tilde{i}}$ is the total probability of making a transition between them.

Since the time-reversal operator and the Hamiltonian commute ($[\hat{T}, \hat{H}] = 0$), it follows that $H_{ji'} = H_{j'i}$, and since $\hat{T}^2 = 1$, $H_{j'i'} = H_{ji}$. Then, using expressions for $|\tilde{i}\rangle$ and $|\tilde{j}\rangle$ from Eq. (4.42) we get

$$\tilde{H}_{\tilde{j}\tilde{i}} = \frac{H_{ji} + zH_{j'i}}{N_i N_j}. \tag{4.46}$$

To evaluate $P_{\tilde{j}\tilde{i}}$, treat $|i\rangle_{\text{rep}}$ just as the usual (unsymmetrized) state (only consider the usual Hamiltonian connections from $|i\rangle_{\text{rep}}$, not its symmetry related pair state). With probability $P_{j,i_{\text{rep}}}$, state $|i\rangle_{\text{rep}}$ spawns to state $|j\rangle$, which may or may

not be its own representative. Since non-representative states are not included in the list of occupied states, if $|j'\rangle \neq |j\rangle$, we must sum over both possibilities to get

$$\begin{aligned} P_{\tilde{j}\tilde{i}} &= P_{j_{\rm rep},i_{\rm rep}} \\ &= P_{j,i_{\rm rep}} + P_{j',i_{\rm rep}}. \end{aligned} \tag{4.47}$$

4.F Choice of weighting when multiple events lead to a single state

Let P be the matrix element of the projector for a transition between two particular states. Suppose there are N events with probabilities $p_i, \cdots, p_N$ that result in this transition. (In general, $\sum_i^N p_i < 1$, since there are other states that can be accessed also.) The probability of getting that state is $\sum_i^N p_i$, so the weight multiplier for this move is $\frac{P}{\sum_i^N p_i}$ regardless of which event led to this state. This prescription is correct since

$$\sum_i^N p_i \, \frac{P}{\sum_i^N p_i} = P. \tag{4.48}$$

(P is independent of i because all the possibilities, i, correspond to the same state.)

There is an alternative approach which avoids the expense of computing the probabilities for the $N-1$ other events that result in the same state. In this approach, when we pick a particular event, i, the weight multiplier is $\frac{P}{Np_i}$ rather than $\frac{P}{\sum_i^N p_i}$, so it depends on which of the N events was selected. This prescription is also correct since

$$\sum_i^N p_i \frac{P}{Np_i} \;=\; P. \tag{4.49}$$

There is a loss of efficiency, that can be large if the p_i differ greatly from each other. So, there is a trade-off between avoiding the expense of computing the $N-1$ additional probabilities and the increase in fluctuations of the weights.

Figure 4.4: Alias method setup

```
! Inputs:  M = number of discrete states to sample
!          {P(i)} = discrete set of probabilities of sampling state i
!
! Outputs: {A(i)} = aliases of states
!          {Q(i)} = probabilities of returning state i rather than its
!                   alias A(i)

! Scale probabilities by M and place in two lists, those smaller and
! those bigger than 1.
n_s = 0 ; n_b = 0
do i=1,M
  A(i) = i
  Q(i) = M*P(i)
  if (Q(i)<1) then
    n_s = n_s + 1
    smaller(n_s) = i
  else
    n_b = n_b + 1
    bigger(n_b) = i
  endif
enddo

! For each orbital construct probability of staying on the orbital and
! the alias of the orbital.
do while (n_s > 0 .and. n_b > 0)
  s = smaller(n_s)
  b = bigger(n_b)
  A(s) = b
  Q(b) = Q(b) + Q(s) - 1
  if (Q(b) < 1) then
    smaller(n_s) = b
    n_b = n_b - 1
  else
    n_s = n_s - 1
  endif
enddo
```

CHAPTER 5

HEAT-BATH CONFIGURATION INTERACTION

I now introduce a new selected configuration interaction algorithm that is based on a deterministic analog of the Efficient Heat-bath Sampling algorithm from Chapter 4. This Heat-bath Configuration Interaction (HCI) algorithm makes use of two parameters, one which controls the process of selecting determinants to add to a variational wavefunction, and one which controls the speed/accuracy tradeoff in computing the perturbative correction to the variational energy. I show that HCI provides a balanced treatment of static and dynamic correlation by computing the potential energy curve of the multireference carbon dimer. I then demonstrate the speed and accuracy of HCI by recovering the full configuration interaction energy of the strongly-correlated chromium dimer in an SVP basis, correlating all 48 electrons, to an accuracy of less than 1 mHa in just a few minutes on a single core. This chapter was adapted from my recent paper published in the Journal for Chemical Theory and Computation [24].

5.1 Introduction

The search for a general, accurate and efficient algorithm for finding approximate solutions to the quantum many-body problem is one of the major open problems in electronic structure theory. Methods that work in the basis of Slater determinants are particularly convenient, because they incorporate the anti-symmetry of the wavefunction, and classes of one-particle basis sets have been developed that enable smooth extrapolation to the complete basis set limit. However, the exponential scaling of the size of the Hilbert space with system size limits deterministic Full Configuration Interaction (Full CI) calculations to very small systems ($\sim 10^{10}$ determinants on a single core).

Stochastic [12, 9, 39] and semistochastic [31] algorithms have shown great promise in sampling the Full CI ground and excited states. Deterministic, variational methods based on tensor networks, such as the Density Matrix Renormalization Group (DMRG) [44, 15], are also routinely applied to strongly-correlated systems. However, (semi)stochastic algorithms have to introduce an approximation such as the *initiator* approximation [18] to overcome the infamous Fermion sign problem, while computationally tractable tensor network states, such as matrix product states and tree tensor network states [29], are inefficient in describing entanglement in molecules that are not quasi-one-dimensional or tree-like, respectively.

Another well-known approach is selected configuration interaction plus perturbation theory (SCI+PT) algorithms [25, 14, 33, 41], which aim to solve the quantum many-body problem within a selected set of determinants. The first SCI+PT method, known as Configuration Interaction by Perturbatively Selecting Iteratively (CIPSI) [25], iteratively augments a selected space of determinants as

follows. At each iteration, the ground state within the selected space is obtained, and the most important components in the first-order perturbation theory correction to that wavefunction are added to the selected space. After convergence or computational limits are reached, second-order perturbation theory is performed on the final variational wavefunction to estimate the Full CI energy. A recent SCI+PT algorithm called Adaptive Sampling CI (ASCI) [41] accelerates CIPSI by generating connections from only those determinants that have a sufficiently large amplitude rather than all the determinants in the current variational wavefunction. In fact the same method had previously been used by two of us for generating trial wavefunctions and deterministic subspaces in the Semistochastic Full Configuration Interaction Quantum Monte Carlo (S-FCIQMC) algorithm [31, 23]. ASCI has contributed to renewed interest in SCI+PT methods by reproducing frozen-core DMRG energies for the challenging Cr_2 molecule to within 0.6 mHa in a few CPU hours, albeit in a very small basis, and provided the motivation for this work.

While CIPSI, ASCI, and other SCI+PT algorithms greatly reduce the number of determinants included in the variational wavefunction, they nevertheless are computationally demanding. The reason for this is that both SCI+PT steps − identifying new determinants to add to the selected space and computing a perturbation theory correction to the energy of the variational wavefunction − require examining *all* of the determinants connected to a reference determinant by nonzero Hamiltonian matrix elements. However, as can be seen in Fig. 5.1, the distribution of magnitudes of off-diagonal matrix elements connected to a reference can span many orders of magnitude. Determinants that are connected to the reference by very small Hamiltonian matrix elements will not be added to the space of the variational wavefunction, and will contribute little to the perturbative correction. In addition, the coefficients of the reference determinants also vary

by orders of magnitude, and the identification of new determinants to include depends on the product of these coefficients and the matrix elements. The method we present scans only those determinants for which the above product is above some threshold, thereby greatly reducing the computer time.

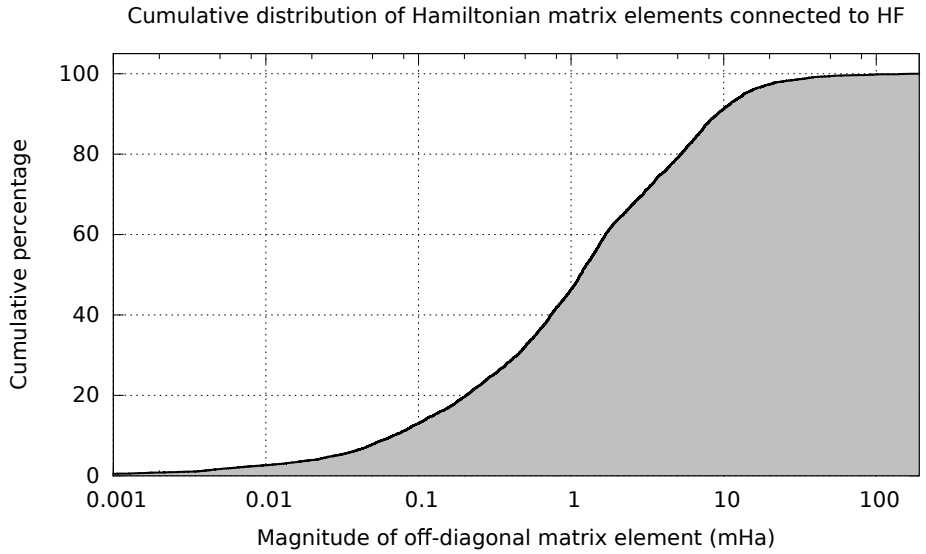

Figure 5.1: Cumulative distribution of the magnitudes of off-diagonal Hamiltonian matrix elements connected to the Hartree-Fock determinant for Cr_2 at $r = 1.5$ Å, in the Ahlrichs VDZ basis [32], where all 48 electrons were correlated. Hartree-Fock orbitals were used. All 30 298 double-excitation matrix elements larger in magnitude than 10^{-8} Ha were included in the cumulative probabilities. Whereas the largest-magnitude off-diagonal matrix element is 191 mHa, about 97% of the matrix elements are at least 10 times smaller (i.e., smaller than 19.1 mHa), 63% are at least 100 times smaller, and some are even more than 1 000 000 times smaller than the largest. Therefore, generating *a*ll determinants connected to a reference is an inefficient use of computational resources. It is more efficient to generate only those determinants that are connected to the reference by matrix elements larger than a threshold, as described in the text.

To solve this problem, in this chapter, we first introduce a new algorithm for generating determinants connected to a reference determinant, which we re-

fer to as deterministic heat-bath sampling because it is a deterministic analog of the (stochastic) efficient heat-bath sampling algorithm two of us recently developed [23]. Instead of generating *all* single and double excitations, as is typically done in quantum chemistry algorithms, we instead generate only those single and double excitations corresponding to Hamiltonian matrix elements exceeding a threshold ϵ. The time complexity of our algorithm scales only as the number of determinants that meet the cutoff; no time is wasted on double excitations that do not.

We then incorporate this deterministic heat-bath algorithm into a new SCI+PT algorithm, which we call Heat-bath CI (HCI), in which both the selection of new determinants for the variational wavefunction and the computation of the perturbative correction are greatly accelerated by skipping over determinants connected by small matrix elements. HCI is capable of quickly and accurately describing electron correlation.

This chapter is organized as follows. In section 5.2, we describe our deterministic heat-bath algorithm for generating only the important determinants connected to a reference determinant. In section 5.3, we describe HCI, our new SCI+PT algorithm, which is based on our deterministic heat-bath algorithm. In section 5.4, we apply HCI to the potential energy curve of the carbon dimer and to the chromium dimer, to benchmark its accuracy against highly accurate algorithms such as Full CI and DMRG. Finally, in section 5.5 we describe our current research directions.

5.2 Deterministic Heat-Bath "Sampling"

Here we describe our algorithm for generating only those determinants that are connected to a single reference determinant by Hamiltonian matrix elements whose magnitudes exceed a cutoff ϵ. We call this a deterministic heat-bath algorithm because it is the deterministic analog of the efficient heat-bath sampling algorithm recently developed by two of us [23]. Stochastic efficient heat-bath sampling enables one to efficiently sample the determinants connected to a reference determinant according to an approximate heat-bath distribution, i.e., with probability approximately proportional to the absolute value of the Hamiltonian matrix element connecting the target determinant to the reference determinant.

While configuration state functions (CSFs) and non-orthogonal Slater determinants certainly have their advantages [13, 21, 40], we instead choose to work in the space of orthogonal determinants because the quantum chemical Hamiltonian takes on a very simple structure [26], which we use to our advantage. The inspiration for this deterministic heat-bath algorithm comes from two observations on the form of the quantum chemical Hamiltonian:

1. Most of the nonzero off-diagonal elements of the Hamiltonian are double excitations.
2. The magnitudes of double excitations (not the signs) depend only on the four orbitals that change occupancy during the excitation and not on any other orbitals.

Because of observation 1, we will get the most gain by accelerating the algorithm for generating double excitations. Because of observation 2, we can organize and

store the magnitudes of double excitations any way we choose once and for all at the beginning of a run.

Let $|H(rs \leftarrow pq)|$ denote the magnitude of a double excitation in which electrons in spin-orbitals p and q excite to spin-orbitals r and s. As noted in observation 2, the magnitude of this matrix element depends only on the spin-orbitals $\{p, q, r, s\}$, independent of which other spin-orbitals are occupied.

The deterministic heat-bath algorithm has two parts: a setup routine, to be called once at the beginning of the run, and a routine for generating determinants connected to a reference determinant by matrix elements with magnitude larger than a threshold, that can be called any time the important determinants connected to a reference are needed. This latter routine will be a crucial component of our Heat-bath CI algorithm, described in section 5.3.

5.2.1 Setup

Store the set of double excitations as follows: For each pair of orbitals $\{p, q\}$, store a list of triplets $\{r, s, |H(rs \leftarrow pq)|\}$, one triplet for each distinct pair of orbitals $\{r, s\}$ that do not include $\{p, q\}$, sorted by $|H(rs \leftarrow pq)|$ in decreasing order. Also, store the maximum magnitude of a double excitation, $H_{\max}^{\text{doub}}$.

This setup has time complexity $\mathcal{O}(M^4 \log M)$ and space complexity $\mathcal{O}(M^4)$, where M is the number of orbitals.

5.2.2 Generating connected determinants

When generating determinants connected to a reference determinant, the usual strategy employed by other quantum chemistry methods such as CISD, MP2, CIPSI and ASCI is to generate *all* single and double excitations from the reference determinant. However, we do not do this because many of the matrix elements connecting these excited determinants to the reference determinant are very small, and the time required to generate those excitations may be better spent elsewhere.

Instead, we introduce a threshold, ϵ, and generate only those determinants that are connected to the reference determinant by Hamiltonian matrix elements that are larger in magnitude than ϵ, as follows:

1. **Generate only those double excitations that exceed ϵ.** If $\epsilon > H_{\text{max}}^{\text{doub}}$, no double excitations are generated. Otherwise, loop over all pairs of occupied orbitals $\{p, q\}$. For each pair, look up the stored list of triplets $\{r, s, |H(rs \leftarrow pq)|\}$, omitting those in which r or s is occupied, until a triplet is reached for which $|H(rs \leftarrow pq)| < \epsilon$.

2. **Generate all single excitations, then discard those that are smaller than ϵ.** Loop over all occupied orbitals p. For each p, loop over all orbitals $\{r\}$ in the same irreducible representation, omitting those in which r is occupied in the reference determinant. Compute $H(r \leftarrow p)$, the matrix element corresponding to the single excitation in which an electron moves from orbital p to orbital r. If $|H(r \leftarrow p)| < \epsilon$, discard the single excitation $(r \leftarrow p)$. The cost of generating the single excitations is $\mathcal{O}(N^2 M)$, where N and M are the numbers of electrons and orbitals, respectively, since each single excitation matrix element takes $\mathcal{O}(N)$ time to compute.

This algorithm has a time complexity of $\mathcal{O}\left(N_{\text{con}}^{\epsilon} + N^2 M\right)$, where $N_{\text{con}}^{\epsilon}$ is the number of determinants connected to the reference by matrix elements that are larger in magnitude than ϵ. No time is wasted on those doubly-excited determinants connected to the reference by matrix elements smaller in magnitude than ϵ.

By varying the threshold ϵ, one can vary between "accurate but slow" (small ϵ) and "less accurate but fast" (large ϵ), depending on the demands of the system being studied. When $\epsilon = 0$, this algorithm reduces to the standard algorithm of generating *all* determinants connected to a reference, i.e., "accurate but probably unnecessarily slow."

5.3 Heat-bath Configuration Interaction

In this section, we apply the deterministic heat-bath sampling algorithm of section 5.2 to invent an SCI+PT algorithm, which we call Heat-bath CI (HCI). There are two stages: generating the variational wavefunction and energy, and computing the perturbative energy correction. We formulate our algorithm such that it has only two parameters, one for each stage.

5.3.1 Generating the variational wavefunction

Like CIPSI and ASCI, HCI generates a variational wavefunction using an iterative process in which at each iteration we diagonalize in the selected space, and then add new determinants to the space. In order to identify the new determinants $\{D_k\}$ to add, both CIPSI and HCI choose those determinants that are most "important" according to some importance measure $f\left(D_k\right)$, from among the set of determinants

connected to the current selected space by nonzero Hamiltonian matrix elements. (ASCI, in common with S-FCIQMC [31, 23], chooses new determinants from only those connected to a truncated subspace of the current selected space, but is otherwise identical to CIPSI.) However, the importance measures used by CIPSI and HCI are different. CIPSI uses the first-order perturbation theory estimate of the coefficients, i.e.,

$$f_{\text{CIPSI}}\left(D_k\right) = \left|c_k^{(1)}\right| \quad = \quad \left|\frac{\sum_i H_{ki} c_i}{E_0 - H_{kk}}\right|, \tag{5.1}$$

whereas HCI uses the simpler measure,

$$f_{\text{HCI}}\left(D_k\right) \quad = \quad \max_i \left(\left|H_{ki} c_i\right|\right). \tag{5.2}$$

which completely eliminates the need to query unimportant determinants, as discussed shortly.

This measure is justified as follows. As previously demonstrated in Fig. 5.1, for fixed i, the range of values that $|H_{ki}|$ can take spans many orders of magnitude. The range of possible values of $|H_{ki}c_i|$ is even larger, since the coefficients $\{c_i\}$ can also vary widely in magnitude. On the other hand, the denominators in Eq. 5.1 do not vary as widely, and are unlikely to range in value by even one order of magnitude. This is particularly true in later iterations, since by then, determinants with low diagonal matrix elements H_{kk} have already been included in the wavefunction. Therefore, most of the variation in Eq. 5.1 is dominated by $\max_i \left(\left|H_{ki} c_i\right|\right)$.

Because the variations in both importance measures are dominated by variations in $\max_i \left(\left|H_{ki} c_i\right|\right)$, they are likely to yield very similar rankings of the candidate determinants by importance. Recall that both CIPSI and HCI use their respective importance measures only to divide the determinants into two groups (important ones to add to the selected space, and unimportant ones to discard).

Small differences in ordering make no difference unless they cause determinants to move between the two groups, and those that switch groups are likely to carry a small weight after diagonalization anyway. Fig. 5.2 shows a comparison of two 5-iteration runs, in which the same numbers of determinants were added each iteration, but with the new determinants selected according to the two different importance measures. As expected, the two importance measures produce variational wavefunctions that are very similar in energy on each iteration.

This new measure for selecting determinants (Eq. 5.2) enables the computational cost of HCI to be much less than that of CIPSI for two reasons. First, we examine only those determinants that are connected to a determinant in the reference wavefunction by Hamiltonian matrix elements larger than a threshold, which avoids wasting time on determinants that are unlikely to be important components of the first-order correction. Second, since all of these determinants are connected strongly enough to the reference to have already met a threshold, we add all of them to the selected space, and bypass the costly computation of first-order perturbation theory estimates of the coefficients altogether.

To accomplish this, we use our deterministic heat-bath sampling algorithm, described in section 5.2. This stage of the algorithm requires one parameter, ϵ_1. On the first iteration, we start with a selected space consisting of only the Hartree-Fock determinant. We iterate as follows.

1. Find the lowest eigenvector of the Hamiltonian in the selected space, and denote the determinant coefficients as $\{c_i\}$.
2. Find all determinants $\{D_k\}$ outside of the selected space for which $|H_{ki}c_i| > \epsilon_1$ for at least one determinant D_i in the selected space.
3. Add those determinants to the selected space.

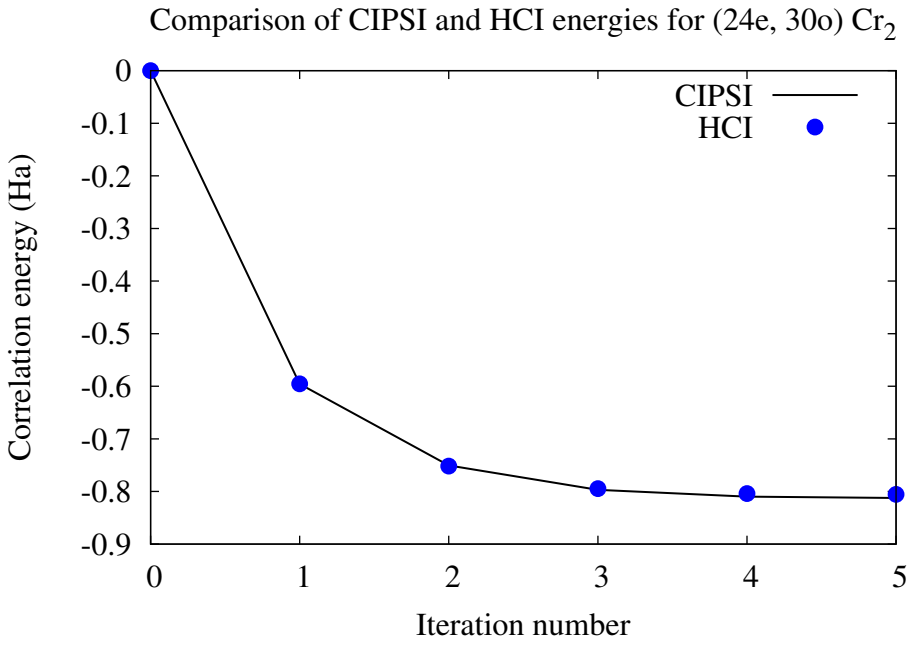

Figure 5.2: Comparison of the variational correlation energies obtained using the CIPSI and HCI importance measures (Eqs. 5.1 and 5.2), for the chromium dimer at $r = 1.5$ Å, in the Ahlrichs VDZ basis [32]. Natural orbitals from a CASSCF(12,12) were used, and Mg cores were frozen. The HCI energies were obtained using $\epsilon_1 = 1$ mHa, and it converged in 5 iterations. The CIPSI energies were obtained by adding the same number of determinants that HCI added each iteration, but the new determinants were chosen by searching all determinants connected to the previous iteration's ground state and choosing those with the largest importance according to Eq. 5.1. Although the two importance measures, f_{CIPSI} and f_{HCI}, are different, the variational wavefunctions they produce iteratively are very similar in energy. The variational energy from HCI is 7 mHa higher at the final iteration. After the perturbative correction the difference in the energies is much smaller than this and can be of either sign.

4. Repeat steps 1-3 until an iteration is reached in which the number of new determinants is less than 1% of the number of determinants already selected.

Step 2 can be accomplished efficiently as follows: Iterate over all determinants $\{D_i\}$ in the selected space. For each, use the deterministic heat-bath sampling algorithm to generate all determinants connected to D_i by Hamiltonian matrix elements larger than $\epsilon = \epsilon_1 / |c_i|$. Then, the lists of connected determinants are merged (duplicates are removed).

5.3.2 Perturbative correction

The second-order perturbative correction to the variational energy $E^{(0)}$ is given by

$$\Delta E^{(2)} = \sum_k \frac{\left(\sum_i H_{ki} c_i\right)^2}{E^{(0)} - H_{kk}}, \tag{5.3}$$

where the sum on k runs over all determinants outside of the selected space, that are connected to at least one determinant in the selected space by a nonzero Hamiltonian matrix element.

As previously mentioned, the magnitudes of $\{H_{ki}\}$ can span many orders of magnitude, so many terms in the sum above are very small. We therefore introduce an approximate perturbation theory expression that makes use of one more parameter, ϵ_2:

$$\Delta E^{(2)} \approx \sum_k \frac{\left(\sum_i^{(\epsilon_2)} H_{ki} c_i\right)^2}{E^{(0)} - H_{kk}}, \tag{5.4}$$

where $\sum^{(\epsilon_2)}$ denotes a sum in which all terms in the sum smaller in magnitude than ϵ_2 are removed. In other words, we approximate the sum on i in the numerator by skipping over the small contributions $H_{ki} c_i$ for which $|H_{ki} c_i| < \epsilon_2$.

This stage can be accomplished efficiently using the deterministic heat-bath algorithm of Section 5.2 as follows. Iterate over all determinants $\{D_i\}$ in the variational wavefunction. For each one, generate all determinants $\{D_k\}$ connected to D_i by matrix elements larger than $\epsilon = \epsilon_2 / |c_i|$, and store both their labels and corresponding values of $H_{ki}c_i$. Once all have been generated, merge the lists, so that the sums $\sum_i H_{ki}c_i$ can be computed before they are squared.

Note that this naïve approach requires storing all of the determinants that contribute to the second-order energy, so it will be too expensive when that number is large. We are currently developing an alternative method with a much smaller storage requirement, which will enable us to use much larger variational determinant expansions. Nevertheless, this simple approach suffices to obtain all the results in this chapter in a few minutes on a single core.

By varying ϵ_2, we can obtain approximate values for the second-order energy correction quickly, without wasting any time on the small terms that are excluded from the sum. We can also extrapolate to the limit $\epsilon_2 \to 0$, even when computing the exact expression ($\epsilon_2 = 0$) would be prohibitively expensive.

5.3.3 Context within existing quantum chemistry algorithms

We now place HCI in the context of other quantum chemistry algorithms. An HCI run is specified by two parameters: ϵ_1, which controls which determinants will be included in the variational wavefunction, and ϵ_2, which controls the accuracy of the perturbative correction to the variational energy. We therefore denote a particular instance of HCI as HCI(ϵ_1,ϵ_2).

When ϵ_1 is larger than the magnitudes of all off-diagonal Hamiltonian matrix elements connected to the Hartree-Fock (HF) reference, no determinants will be added to the variational wavefunction. In that case, the variational wavefunction will reduce to the HF determinant. When the variational wavefunction is HF, the perturbative correction will yield the second-order Epstein-Nesbet perturbation theory (EN-PT) energy correction if $\epsilon_2 = 0$, and will yield zero if ϵ_2 is larger than the magnitudes of all off-diagonal Hamiltonian matrix elements connected to HF (since no terms would be included in the sum).

When $\epsilon_1 = 0$, all determinants will (after many iterations) be added to the selected space, resulting in a variational wavefunction equal to the Full CI (FCI) ground state. In that case, the perturbative correction will be zero, no matter what value ϵ_2 has, since there will be no other determinants left.

In summary,

$$\mathrm{HCI}\,(\epsilon_1, \epsilon_2) \quad = \quad \begin{cases} \text{HF}, & \text{if} \quad \epsilon_1 \geq H_{\max}^{\mathrm{doub}} \quad \text{and} \quad \epsilon_2 \geq H_{\max}^{\mathrm{doub}}; \\ \text{EN-PT}, & \text{if} \quad \epsilon_1 \geq H_{\max}^{\mathrm{doub}} \quad \text{and} \quad \epsilon_2 = 0; \\ \text{FCI}, & \text{if} \quad \epsilon_1 = 0. \end{cases} \tag{5.5}$$

HCI can thus be seen as a generalization of HF, EN-PT, and FCI, that is more flexible than any of them in enabling a tradeoff between the speed and accuracy of its ground state energy calculations. As we shall see in the following section, FCI-quality results can be obtained at a significantly reduced cost by choosing ϵ_1 and ϵ_2 appropriately.

5.4 Results

We applied our HCI algorithm to two systems, the carbon dimer and the chromium dimer. All integrals, coupled cluster calculations, and orbital rotations were computed using the MOLPRO quantum chemistry software package [43].

5.4.1 Carbon dimer

We first applied HCI to the potential energy surface of the carbon dimer, a system known to be of highly multireference character, even at equilibrium geometry. We used the cc-pVDZ basis set [20] (28 spatial orbitals) and correlated all 12 electrons. The Full CI space consists of about 2×10^{10} Slater determinants. We compared our results to recently-published Full CI values [34], since they were tabulated for a large part of the binding curve. By running tests at equilibrium geometry at various values of ϵ_1 and ϵ_2, we found that using $\epsilon_1 = 1$ mHa and $\epsilon_2 = 30$ μHa gave a ground-state energy converged to within 1 mHa of the Full CI energy. We then used these parameter values for the whole curve.

We compared to the "gold standard" of quantum chemistry, coupled cluster with singles, doubles, and perturbative triples (CCSD(T)). As can be seen in Fig. 5.3, HCI describes the whole binding curve well, whereas CCSD(T) is only accurate near equilibrium. This shows that HCI can capture both static and dynamic correlation effects. In Fig. 5.4, we see that, even near equilibrium, where CCSD(T) is supposed to be good, HCI still gets closer to the Full CI energy at these choices of ϵ_1 and ϵ_2.

We used only D_{2h} symmetry, rather than the full $D_{\infty h}$ symmetry, so both

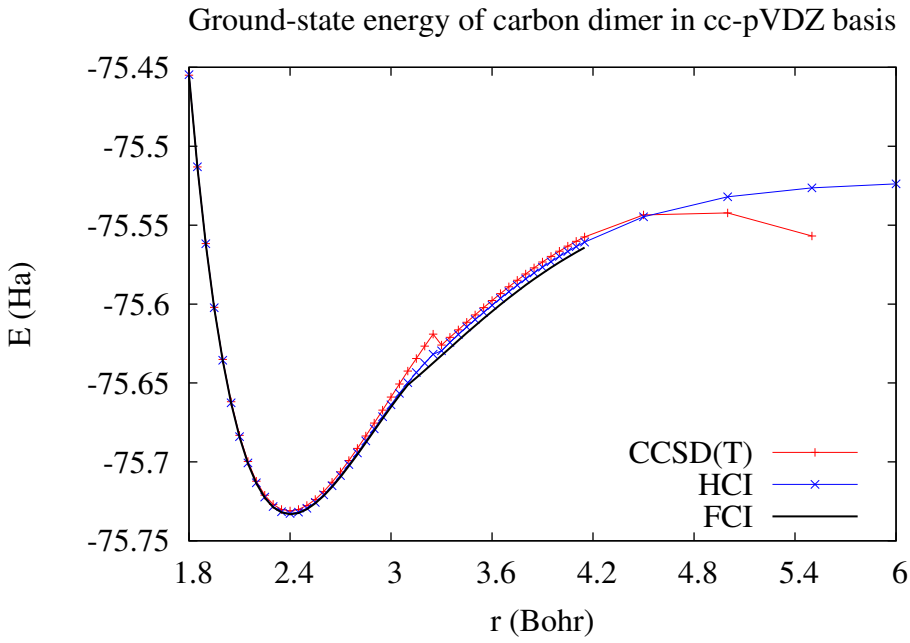

Figure 5.3: Binding curve of C_2 in a cc-pVDZ basis obtained from HCI($\epsilon_1 = 1$ mHa, $\epsilon_2 = 30$ μHa) compared to CCSD(T) (computed using MOLPRO) and Full CI [34]. CCSD(T) is good at describing dynamic correlation but poor at describing static correlation, so while it gives good energies near equilibrium, it can't describe bond breaking well. On the other hand, HCI gives good energies along the whole dissociation curve, indicating that it accurately describes both static and dynamic correlation.

coupled cluster and HCI were less accurate in the region near $r = 3.25$ Bohr, where a curve crossing occurs between curves of different symmetries. This may explain why HCI took 24 ± 4 seconds to compute each energy in the range from $r = 2.75$ to 3.30 Bohr, while it only took 13 ± 1 seconds to compute each energy outside of that region.

In larger basis sets, the rate of convergence with respect to ϵ_1 and ϵ_2 can be slower. For example, in Fig 5.5, we plot the convergence of the ground state energy of the carbon dimer in the cc-pVTZ basis at equilibrium. Several of the energies

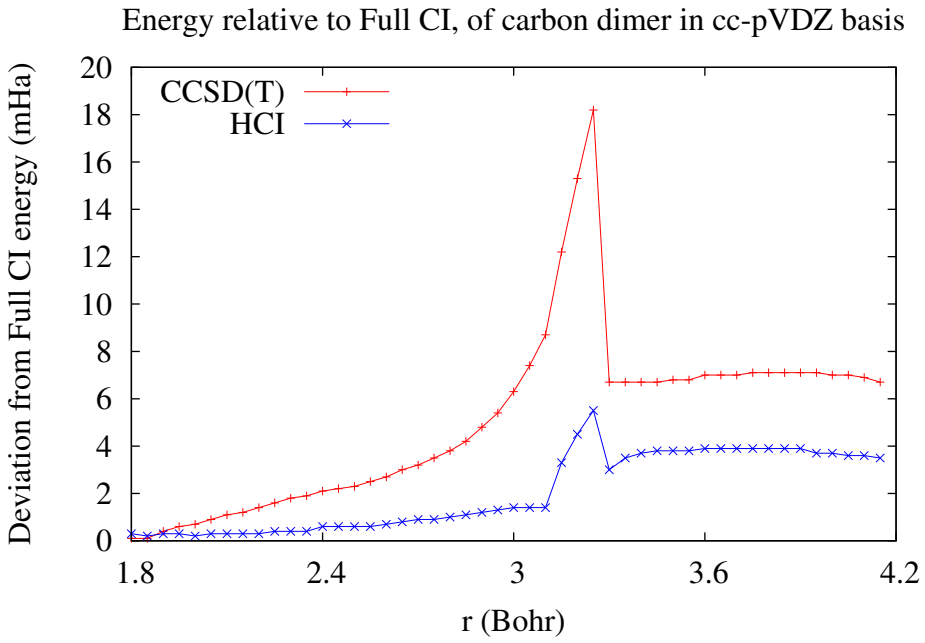

Figure 5.4: Energy deviations relative to Full CI. HCI gets closer to the Full CI energy, even in the region near equilibrium ($r = 2.4$ Bohr). The discontinuity near $r = 3.25$ Bohr is due to a jump in the Hartree-Fock solution due to a curve crossing between curves of different symmetries [34] and the fact that Hartree-Fock is not guaranteed to find the global minimum [1]. Both methods have difficulty in this region, but HCI has much less difficulty, since its variational stage can describe the multireference character of the molecule.

for the smaller ϵ_1 and ϵ_2 values are within 1 mHa of the extrapolated energy and these runs took between 4 and 9 minutes, compared to about 10 seconds for the cc-pVDZ basis.

5.4.2 Chromium dimer

We performed ground-state energy calculations on the challenging Cr_2 dimer in the Ahlrichs VDZ basis [32] at $r = 1.5$ Å, both with frozen Mg core (24e, 30o),

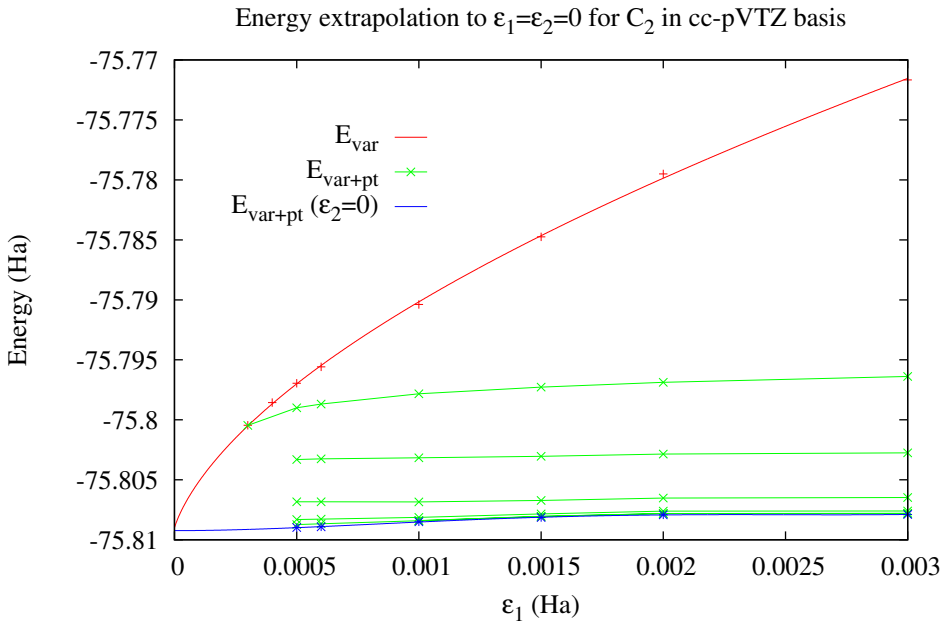

Figure 5.5: Plot of the convergence to the Full CI limit ($\epsilon_1 = \epsilon_2 = 0$) of the ground-state energy of the carbon dimer at the equilibrium bond length of 1.24253 Å in the cc-pVTZ basis set, with all electrons excited. CASSCF(8,8) natural orbitals were used. The Full CI space for this system consists of about 3×10^{14} Slater determinants. The red points and line plot the variational energies, which depend only on ϵ_1. The green points and lines plot total energies (variational + perturbative correction) for 5 different ϵ_2 values ranging from 3 to 300 μHa (two of the lines for the smallest values of ϵ_2 are indistinguishable on the scale of the plot). Finally, the blue points and line plot the total energies for different values of ϵ_1 at $\epsilon_2 = 0$ (i.e., the perturbative correction formula is extrapolated to obtain the exact limit). The extrapolation function for the variational energies was chosen to be a rational function of $\sqrt{\epsilon_1}$. The total energies were extrapolated by first getting the total energy for each ϵ_1 value and $\epsilon_2 = 0$ using a rational polynomial in ϵ_2 shown as blue stars, and then fitting these values to a polynomial in ϵ_1. Our extrapolated ground state energy is -75.80924(15) Ha, in agreement with the value -75.809285 obtained from the DMRG calculation [30] with the largest bond dimension. The lowest computed energy is -75.80873 for $\epsilon_1 = 5$ mHa and $\epsilon_2 = 3$ μHa and the uncertainty in our extrapolated energy is given as 1/2 of the energy extrapolation.

and with all electrons correlated (48e, 42o). The Full CI spaces of these systems are approximately 9×10^{14} and 2×10^{22} determinants, respectively. We compared to extrapolated Density Matrix Renormalization Group (DMRG) results [30].

We used natural orbitals from a (12e, 12o) CASSCF. For both the 24- and 48-electron cases, we used $\epsilon_1 = 1$ mHa and $\epsilon_2 = 10$ μHa. As shown in table 5.1, this choice in parameters enabled HCI to produce energies within 1 mHa of the converged DMRG results.

	System	
Method	(24e, 30o)	(48e, 42o)
HCI	-0.421 30	-0.444 04
DMRG	-0.420 95(3)	-0.444 78(32)

Table 5.1: Energies (E + 2086 in Ha) of Cr_2 in the Ahlrichs VDZ basis at $r = 1.5$ Å, from HCI($\epsilon_1 = 1$ mHa, $\epsilon_2 = 10$ μHa) and converged DMRG [30]. At these values of ϵ_1 and ϵ_2, the energies are within 1 mHa of the converged DMRG results.

At $\epsilon_1 = 1$ mHa, for (24e, 30o) and (48e, 42o) respectively, the variational wavefunctions consisted of 42 945 and 63 592 determinants, with variational energies -2086.368 and -2086.384 Ha.

These HCI runs took only about two and eight minutes for (24e, 30o) and (48e, 42o) respectively, on a single core. This demonstrates that it is not significantly more challenging for HCI to approximate the Full CI energy in this basis with all electrons correlated than with core electrons frozen. We believe that this is because the Hamiltonian matrix elements corresponding to core-valence excitations tend to be small (because of the small overlap between sharp and diffuse orbitals), and HCI efficiently includes only the most important excitations of that type. However, note that whereas this Ahlrichs VDZ basis allows core-valence excitations, it is

not flexible enough to allow for core-core excitations, so this conclusion may not generalize to basis sets such as Dunning's cc-pCVDZ basis [20].

A convergence plot of the ground-state energy with respect to the two input parameters ϵ_1 and ϵ_2 is given in Fig. 5.6. In the limit that $\epsilon_1 = \epsilon_2 = 0$, the Full CI energy is obtained, but highly-accurate results can be obtained much more cheaply at small but nonzero values of ϵ_1 and ϵ_2. Although the perturbative estimate is an underestimate of the true correction in Fig. 5.5 and an overestimate in Fig. 5.6, it is apparent from both figures that as the variational wavefunction improves, the perturbative correction becomes progressively more effective at recovering the missing energy.

For the converged runs that we did for C_2 in cc-pVDZ basis and Cr_2 in Ahlrichs VDZ basis, the variational and perturbative stages both took approximately the same amount of time. However, as previously mentioned, the perturbative stage was done in a naïve way that is space-limited.

5.5 Outlook and ongoing research

We believe that the Heat-bath Configuration Interaction (HCI) algorithm described here is an accurate and efficient method for treating many quantum many-body systems, including several that are commonly considered to be strongly correlated. There are many problems where the entire state space is enormous, but the portion of it that makes a significant contribution to the exact many-body wavefunction is sufficiently small that it can be included in the HCI wavefunction. Since varying ϵ_1 and ϵ_2 enables different choices in the speed/accuracy tradeoff, we believe that HCI will be competitive with many other electronic structure methods,

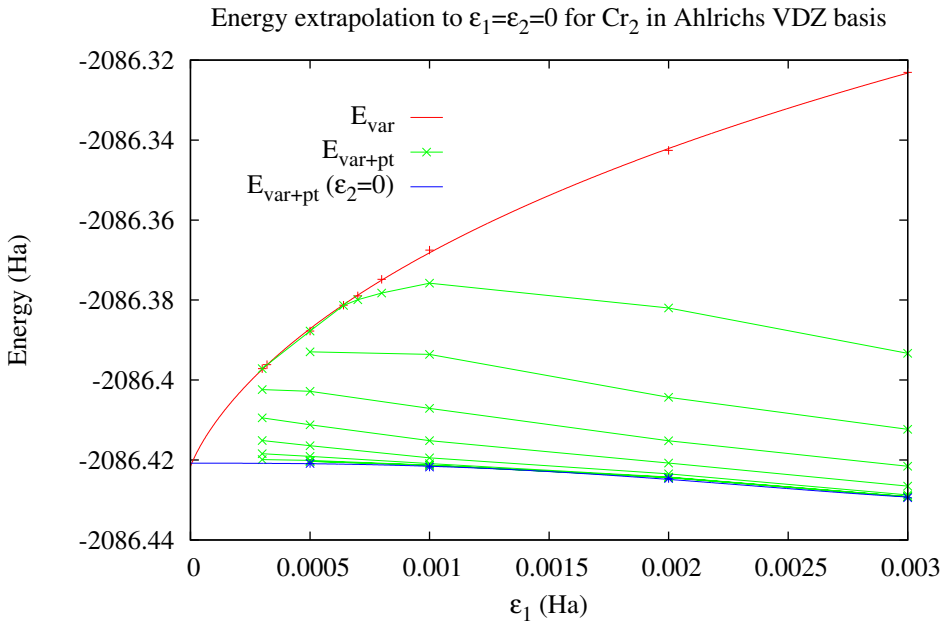

Figure 5.6: Plot of the convergence of the ground-state energy of (24e, 30o) Cr_2 to the Full CI limit ($\epsilon_1 = \epsilon_2 = 0$). The red and blue points and lines have the same meaning as in Figure 5.5. The green points and lines plot total energies (variational + perturbative correction) for 9 different ϵ_2 values ranging from 2.5 to 640 μHa, increasing by factors of 2 as one goes up the graph (several of the lines for the smallest values of ϵ_2 are indistinguishable). Although the energies given in Table 5.1 were not extrapolated, this plot shows how both the variational and total energies can be extrapolated to the Full CI limit. The extrapolation functions were chosen to be of the same functional form as in Fig. 5.5. Our extrapolated ground state energy is -2086.4208(2) Ha, consistent with the extrapolated DMRG energy given in Table 5.1. The uncertainty in the extrapolated energy is given as 1/2 of the energy extrapolation relative to the energy of -2086.42033 Ha obtained for $\epsilon_1 = 0.5$ mHa, $\epsilon_2 = 2.5$ μHa. This was the most computationally-intensive run in this extrapolation plot and took about 14 minutes on a single core.

ranging from fast, approximate methods like Density Functional Theory (DFT), to highly-accurate, expensive methods like Density Matrix Renormalization Group (DMRG) and Semistochastic Full Configuration Interaction Quantum Monte Carlo (S-FCIQMC). The most relevant competing method is DMRG. For 3-dimensional systems, the bond dimension of DMRG scales as $N^{2/3}$ so the computational cost scales exponentially in $N^{2/3}$ whereas the computational cost of HCI is exponential in N. Nevertheless, because HCI has a much smaller prefactor, we expect that it will be the method of choice for some strongly correlated systems.

We are currently working on extending HCI to be able to treat strongly-correlated systems in larger basis sets. One promising idea in this vein is to use subsets of the variational wavefunction at a time, chosen either deterministically or stochastically (using the stochastic efficient heat-bath sampling algorithm [23]). This not only removes the storage bottleneck in our current implementation.

Computing the one- and two-body reduced density matrices of the variational wavefunction is a trivial extension, and it will enable us to both perform orbital rotations and compute ground-state molecular properties other than the energy. Time-reversal symmetry and L_z symmetry for linear molecules can also be used to reduce the effective Hilbert space size. Extension to low-lying excited states is straight-forward using a method analogous to that outlined in the original CIPSI paper [25]. Finally, highly-accurate geometry optimizations can be performed, where early iterations use fast, coarse estimates of the energy (large ϵ_1, ϵ_2), and later iterations fine-tune those geometry configurations with more expensive runs (small ϵ_1, ϵ_2).

We are developing these extensions now.

Acknowledgments: We thank Garnet Chan and Sandeep Sharma for valuable discussions and Sandeep Sharma for help with using MOLPRO and understanding Cr_2 benchmarks and basis sets. This work was supported by grant NSF ACI-1534965. NT was supported through the Scientific Discovery through Advanced Computing (SciDAC) program funded by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research, and Basic Energy Sciences.

CHAPTER 6

CONCLUSION AND OUTLOOK

In this dissertation, I have presented three main electronic structure algorithms, which have achieved an increase in computational efficiency over state-of-the-art algorithms by many orders of magnitude.

The first algorithm, Semistochastic Quantum Monte Carlo, combined the best qualities of deterministic and stochastic projection algorithms, resulting in a mostly-stochastic algorithm in which the most important contribution to the wavefunction is computed deterministically. If the deterministic subspace of the full Hilbert space is well chosen, this corresponds to both a dramatic increase in computational efficiency, and a decrease in *initiator* bias compared the fully-stochastic Full Configuration Interaction Quantum Monte Carlo. Aside from the introduction of a deterministic subspace, we introduced two other improvements: using a multi-determinant trial wavefunction for computing the mixed energy estimator, rather than just the Hartree-Fock determinant, and describing the sparse "snapshots" by random walkers carrying real, rather than integer, weights.

Next, I developed an efficient algorithm for sampling Slater determinants according to an approximate heat-bath distribution. This was achieved by factoring the heat-bath distribution, then introducing an approximation that enabled these approximate factors to be precomputed and stored before the start of a long run. The alias method then enabled sampling of this approximate heat-bath distribution in $\mathcal{O}(N_{\mathrm{el}})$ time, much faster than the $\mathcal{O}(N_{\mathrm{el}}^2 N_{\mathrm{orb}}^2)$ time required to sample the exact heat-bath distribution. This new sampling distribution improves the efficiency of Full Configuration Interaction Quantum Monte Carlo in two ways. First, the fluctuations in walker weights are decreased because newly-spawned walkers

now carry nearly the same weight, regardless of what new determinant they appear on. Second, because the fluctuations are smaller, a much larger time step can now be used. Previously, for interesting systems, a tiny time step had to be used to curb the effects of the giant fluctuations coming from uniform sampling, resulting in very computationally demanding calculations.

Finally, I developed the deterministic analog of my Efficient Heat-bath Sampling algorithm. Rather than efficiently sampling determinants with probability proportional to the corresponding Hamiltonian matrix element, the deterministic version generates all determinants whose corresponding matrix elements exceed a cutoff. This algorithm represents an important development in quantum chemistry, as all previous determinant-based algorithms relied on naively generating all single and double excitations from a reference, whereas my new algorithm only generates the important ones. I then built this deterministic "sampling" algorithm into a Selected Configuration Interaction and Perturbation Theory algorithm I call Heat-bath Configuration Interaction. It dramatically improves upon the performance of related algorithms such as Configuration Interaction by Perturbatively Selecting Iteratively (CIPSI) by not wasting any time on unimportant determinants that are unlikely to be added to the variational wavefunction or are unlikely to contribute significantly to the perturbative correction.

BIBLIOGRAPHY

[1] pages We used the D_{2h} subgroup of the full $D_{\infty h}$ symmetry group for the two–electron integrals. Therefore the $^1\Sigma_g^+$ and $^1\Delta_g$ states of the $D_{\infty h}$ group can mix since they transform as the same irreducible representation of the subgroup. Hartree–Fock (HF) is not guaranteed to find the global minimum–energy determinant, and MOLPRO switches the occupied orbitals and the symmetry of the HF determinant from $^1\Sigma_g^+$ to $^1\Delta_g$ near $r = 3.30$ Bohr if the HF calculations at each distance are done without providing input orbitals (although the $^1\Delta_g$ is the lower energy HF state for $r > 2.20$ Bohr). This results in a discontinuity in the Hartree–Fock energy, and smaller discontinuities in the CCSD(T) and HCI energies since these methods use the HF orbitals. Of course in the $\epsilon_1 \to 0$, $\epsilon_2 \to 0$ limit, the HCI energies would not have a discontinuity since the FCI energy is invariant to orbital rotations. Note that if the HF calculation was done using $D_{\infty h}$ symmetry then there would be only a derivative discontinuity at the curve crossing in the HF energy, but there would still be discontinuities in the CCSD(T) and HCI energies because of the discontinuity in the HF orbitals.

[2] To switch to chemist notation, interchange the middle two indices.

[3] For example: J. E. Hirsch, Phys. Rev. B **31**, 4403 (1985); N. Furukawa and M. Imada, J. Phys. Soc. Jpn. **61**, 3331 (1992); H. D. Raedt and M. Frick, Phys. Rep. **231**, 107 (1993); S. Zhang, J. Carlson, and J. E. Gubernatis, Phys. Rev. B **55**, 7464 (1997).

[4] In the initiator approach, a state must have a minimum absolute weight i in order to contribute an off-diagonal move to a state that is not already occupied. This results in a non-variational bias that vanishes in the limit of infinite walker population. In our modification of this approach, $i = cm^p$, where m is the number of moves the walker has made since its last visit to the deterministic space and c and p are constants, chosen to be 1 for the applications in this paper.

[5] NS Blunt, TW Rogers, JS Spencer, and WMC Foulkes. Density-matrix quantum monte carlo method. *Phys. Rev. B*, 89(24):245124, 2014.

[6] NS Blunt, Simon D Smart, JAF Kersten, JS Spencer, George H Booth, and Ali Alavi. Semi-stochastic full configuration interaction quantum monte carlo: Developments and application. *J. Chem. Phys.*, 142(18):184107, 2015.

[7] George H Booth and Garnet Kin-Lic Chan. Communication: Excited states,

dynamic correlation functions and spectral properties from full configuration interaction quantum monte carlo. *J. Chem. Phys.*, 137(19):191102, 2012.

[8] George H Booth, Deidre Cleland, Alex JW Thom, and Ali Alavi. Breaking the carbon dimer: the challenges of multiple bond dissociation with full configuration interaction quantum monte carlo methods. *J. Chem. Phys.*, 135(8):084104, 2011.

[9] George H Booth, Andreas Grüneis, Georg Kresse, and Ali Alavi. Towards an exact description of electronic wavefunctions in real solids. *Nature*, 493(7432):365–370, 2013.

[10] George H Booth, Simon D Smart, and Ali Alavi. Linear-scaling and parallelisable algorithms for stochastic quantum chemistry. *Mol. Phys.*, 112(14):1855–1869, 2014.

[11] George H Booth, Alex J W Thom, and Ali Alavi. Fermion Monte Carlo without fixed nodes: a game of life, death, and annihilation in Slater determinant space. *J. Chem. Phys.*, 131(5):054106, 2009.

[12] George H Booth, Alex JW Thom, and Ali Alavi. Fermion monte carlo without fixed nodes: A game of life, death, and annihilation in slater determinant space. *J. Chem. Phys.*, 131(5):054106, 2009.

[13] Benoît Braïda, Julien Toulouse, Michel Caffarel, and CJ Umrigar. Quantum monte carlo with jastrow-valence-bond wave functions. *J. Chem. Phys.*, 134(8):084108, 2011.

[14] Robert J Buenker and Sigrid D Peyerimhoff. Individualized configuration selection in ci calculations with subsequent energy extrapolation. *Theor. Chim. Acta*, 35(1):33–58, 1974.

[15] Garnet Kin-Lic Chan and Sandeep Sharma. The density matrix renormalization group in quantum chemistry. *Annu. Rev. Phys. Chem.*, 62:465–481, 2011.

[16] Hitesh J. Changlani, Jesse M. Kinder, C. J. Umrigar, and Garnet Kin-Lic Chan. Approximating strongly correlated wave functions with correlator product states. *Phys. Rev. B*, 80(24), Dec 2009.

[17] Deidre Cleland, George H Booth, and Ali Alavi. Communications: Survival of

the fittest: accelerating convergence in full configuration-interaction quantum Monte Carlo. *J. Chem. Phys.*, 132(4):041103, 2010.

[18] Deidre Cleland, George H Booth, and Ali Alavi. Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum monte carlo. *J. Chem. Phys.*, 132(4):041103, 2010.

[19] T.H. Dunning Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90:1007, 1989.

[20] Thom H Dunning Jr. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2):1007–1023, 1989.

[21] Francesco Fracchia, Claudia Filippi, and Claudio Amovilli. Size-extensive wave functions for quantum monte carlo: A linear scaling generalized valence bond approach. *J. Chem. Theory Comput.*, 8(6):1943–1951, 2012.

[22] JC Greer. Monte carlo configuration interaction. *J. Comp. Phys.*, 146(1):181–202, 1998.

[23] Adam A. Holmes, Hitesh J. Changlani, and C. J. Umrigar. Efficient heat-bath sampling in fock space. *J. Chem. Theory Comput.*, 2016.

[24] Adam A. Holmes, Norm M. Tubman, and C. J. Umrigar. Heat-bath configuration interaction: An efficient selected ci algorithm inspired by heat-bath sampling. *Journal of Chemical Theory and Computation*, 2016.

[25] B Huron, JP Malrieu, and P Rancurel. Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions. *J. Chem. Phys.*, 58(12):5745–5759, 1973.

[26] PJ Knowles and NC Handy. A new determinant-based full configuration interaction method. *Chem. Phys. Lett.*, 111(4):315–321, 1984.

[27] Richard A Kronmal and Arthur V Peterson Jr. On the alias method for generating random variables from a discrete distribution. *Amer. Statist.*, 33(4):214–218, 1979.

[28] Fionn D Malone, NS Blunt, James J Shepherd, DKK Lee, JS Spencer, and

WMC Foulkes. Interaction picture density matrix quantum monte carlo. *J. Chem. Phys.*, 143(4):044116, 2015.

[29] Naoki Nakatani and Garnet Kin-Lic Chan. Efficient tree tensor network states (ttns) for quantum chemistry: Generalizations of the density matrix renormalization group algorithm. *J. Chem. Phys.*, 138(13):134113, 2013.

[30] Roberto Olivares-Amaya, Weifeng Hu, Naoki Nakatani, Sandeep Sharma, Jun Yang, and Garnet Kin-Lic Chan. The ab-initio density matrix renormalization group in practice. *J. Chem. Phys.*, 142(3):034102, 2015.

[31] FR Petruzielo, AA Holmes, Hitesh J Changlani, MP Nightingale, and CJ Umrigar. Semistochastic projector monte carlo method. *Phys. Rev. Lett.*, 109(23):230201, 2012.

[32] Ansgar Schäfer, Hans Horn, and Reinhart Ahlrichs. Fully optimized contracted gaussian basis sets for atoms li to kr. *J. Chem. Phys.*, 97(4):2571–2577, 1992.

[33] Jeffrey B Schriber and Francesco A Evangelista. Communication: An adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy. *J. Chem. Phys.*, 144(16):161106, 2016.

[34] Sandeep Sharma and Ali Alavi. Multireference linearized coupled cluster theory for strongly correlated systems using matrix product states. *J. Chem. Phys.*, 143(10):102815, 2015.

[35] Sandeep Sharma and Garnet Kin-Lic Chan. Communication: A flexible multi-reference perturbation theory by minimizing the hylleraas functional with matrix product states. *The Journal of chemical physics*, 141(11):111101, 2014.

[36] James J. Shepherd, Andreas Grueneis, George H. Booth, Georg Kresse, and Ali Alavi. Convergence of many-body wave-function expansions using a plane-wave basis: From homogeneous electron gas to solid state systems. *Phys. Rev. B*, 86(3):035111, 2012.

[37] Simon D. Smart, George H. Booth, and Ali Alavi. unpublished.

[38] Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 1989.

[39] Seiichiro Ten-no. Stochastic determination of effective hamiltonian for the

full configuration interaction solution of quasi-degenerate electronic states. *J. Chem. Phys.*, 138(16):164126, 2013.

[40] Takashi Tsuchimochi and Seiichiro Ten-no. Black-box description of electron correlation with spin-extended configuration interaction model: Implementation and assessment. *J. Chem. Theory Comput.*, 2016.

[41] Norm M Tubman, Joonho Lee, Tyler Y Takeshita, Martin Head-Gordon, and K Birgitta Whaley. A deterministic alternative to the full configuration interaction quantum monte carlo method. *arXiv preprint arXiv:1603.02686*, 2016.

[42] Alastair J Walker. An efficient method for generating discrete random variables with general distributions. *ACM Trans. on Math. Software (TOMS)*, 3(3):253–256, 1977.

[43] Hans-Joachim Werner, Peter J Knowles, Gerald Knizia, Frederick R Manby, and Martin Schütz. Molpro: a general-purpose quantum chemistry program package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2(2):242–253, 2012.

[44] Steven R White and Richard L Martin. Ab initio quantum chemistry using the density matrix renormalization group. *J. Chem. Phys.*, 110(9):4127–4130, 1999.